ROBUSTNESS AND POWER OF PROCEDURES FOR PAIRWISE MULTIPLE COMPARISONS OF REPEATED MEASURES MEANS IN THE SPLIT PLOT DESIGN

By

HYUNCHUL KIM

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1995

ACKNOWLEDGMENTS

I would like to express my sincerest appreciation to Dr. Algina, chairperson of my doctoral committee. I appreciate the other members of my committee, Dr. M. David Miller, Dr. Linda M. Crocker, Dr. A. Agresti, and Dr. R. Webb. I would like to express my heartfelt thanks to my parents, grandmother, my wife Jeein, son Tahyo, and daughter Onhyo for their sacrifices and prayers. I praise the Lord who fed me in the wilderness with manna.

"Blessed are You, Lord of Israel, our Father, forever and ever. Yours, O Lord, is the greatness, the power and glory, the victory and the majesty; for all that is in heaven and in earth is Yours; Yours is the kingdom, O Lord, and You are exalted as head over all. Both riches and honor come from You, and You reign over all. In Your hand is power and might; in Your hand it is to make great and to give strength to all. Now therefore, our God, we thank You and praise Your glorious name (1 Chronicles 29:10b-13)."

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

ABSTRACT v

CHAPTERS

1 INTRODUCTION 1

The Problem 8
Purpose of the Study 9
Significance of the Study 9

2 LITERATURE REVIEW 10

Linear Model for the Split Plot Design with One Between-Subjects Factor 10
Validity Conditions for Univariate F Tests 11
Effects of the Violations of Validity Conditions for Univariate F Tests 12
Robustness against Non-sphericity 12
Robustness against V-C Heteroscedasticity . . . 15
Effect of Nonnormality on Type I error rates. . 18
Effect of Nonnormality on Type II error rates . 21
Use of the Trimmed t Test under Nonnormality. . 22
Pairwise Multiple Comparison Procedures 26
Summary and Implications 36

3 METHODOLOGY 39

Multiple Comparison Procedures 39
Bonferroni Procedures 39
Modified Sequentially Rejective Bonferroni Procedures 41
Welsh's Step-up Procedures 45
Design 47
Distribution of the Generated Data 47
Sphericity of the Common V-C Matrix 48
Degree of Heteroscedasticity of the V-C Matrix. 48
Total Sample Size 50
Sample Size Ratio 50
The Configuration of the Nonnull Means. 51
Simulation Procedure 52
Data Generation 52

Evaluating the Performance of Test Procedures . 57
Summary 57

4 RESULTS AND DISCUSSION 59

Type I Error Rates 60
Completely True Null Hypothesis 60
Effect of T 63
The T x n_1/n_2 and T x n_1/n_2 x $\Sigma_1:\Sigma_2$ effects 63
Partially True Null Hypothesis 69
The T, CM, and T x CM effects 71
The T x D effect 71
Power 76
Any-Pair Power 76
The T, D, and T x D effects 78
The T x n_1/n_2 and T x n_1/n_2 x $\Sigma_1:\Sigma_2$ effects 78
The T x CM x ε effect 84
Average Per-Pair Power 87
The T, T x CM, and T x CM x ε effects . . 89
The T x D effect 93
All-Pairs Power 95
Effect of ε 95
The T, D, and T x D effects 100

5 DISCUSSION AND CONCLUSIONS 102

Findings of Related Research 102
Effect of Non-sphericity 102
Effect of Nonnormality 103
Effect of Trimmed Means 107
Power Comparisons 107
Effect of ε on Type I Error Rates for Shaffer's Test 108
Conclusions 109
Suggestions for Future Research 110

APPENDIX ESTIMATED TYPE I ERROR RATES AND POWER 113

REFERENCES 153

BIOGRAPHICAL SKETCH 158

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ROBUSTNESS AND POWER OF PROCEDURES FOR PAIRWISE MULTIPLE COMPARISONS OF REPEATED MEASURES MEANS IN THE SPLIT PLOT DESIGN

By

Hyunchul Kim

May 1995

Chairperson: James, J. Algina
Major Department: Foundations of Education

Type I error rates and power of 11 procedures for making planned pairwise multiple comparisons of the repeated measures means in a split plot design were investigated. Using simulated data, Type I error rates for completely true null hypotheses and partially true null hypotheses were estimated. In addition, any-pair power, average per-pair power, and all-pairs power were estimated.

Let $\boldsymbol{\mu}_j$ denote the mean vector in the *j*th level of a between-subjects factor. The 11 procedures were designed to make pairwise comparisons of the elements of $\Sigma\boldsymbol{\mu}_j/J$. In each procedure a Welch-type test statistic, designed for use with heterogeneous variance-covariance (v-c) matrices, was used. Three types of procedures were investigated: Bonferroni procedures, modified sequentially rejective Bonferroni procedures, and Welsch step-up procedures. Within each type, procedures using trimmed data and procedures using untrimmed data were investigated.

The effects of six factors were investigated in the study: (a) distribution of the generated data, (b) sphericity of the common v-c matrix, (c) degree of the heterogeneity of the v-c matrices, (d) total sample size, (e) sample size arrangement, and (e) configuration of means.

All 11 procedures controlled Type I error rates reasonably well. However, when the design was unbalanced and v-c matrices were heterogeneous, the Welsch tests on untrimmed data had a slight liberal tendency. Two tests emerged as the most powerful tests: the Welsch test on untrimmed data and the Welsch test on trimmed data. When the data were normal or slightly leptokurtic, the Welsch test on untrimmed data was the most powerful of the eleven tests. This was true for any-pair power, average per-pair power, and all-pairs power. When the data were strongly leptokurtic, the Welsch test on trimmed data was the most powerful test. Again this was true for all three power criteria.

CHAPTER 1
INTRODUCTION

A repeated measures design consists of repeated observations Y_{ik} on the same dependent variable for every subject i = 1, ..., n. In this dissertation, the design of interest is one in which the repeated measures comprise a single within-subjects factor with levels k = 1, ..., K. Let $\mathbf{y}$ denote the (K x 1) random vector that generates the y_{ik}. Observations on $\mathbf{y}$ are assumed to be independently and identically distributed with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. If there is a between-subjects factor, the design is often called a split plot design and the observations Y_{ijk} for subject i = 1, ..., n_j measured at levels k = 1, ..., K of the within-subjects factor are assumed to be independent and identically distributed within each level (j = 1, ..., J) of the between-subjects factor with common mean vector $\boldsymbol{\mu}_j$ and variance-covariance matrix $\boldsymbol{\Sigma}_j$.

Under normality, the validity conditions for the univariate within-subjects F tests have been investigated by Huynh and Feldt (1970) for the split plot design with one between-subjects factor. They showed that conditions on the $\boldsymbol{\Sigma}_j$ that are necessary and sufficient for valid F tests for the within-subjects effects are (a) homogeneity of the variance-covariance (v-c) matrices, for a suitable set of orthonormalized variables, at all levels of the between-subjects factor, and (b) sphericity of the common v-c matrix. These conditions have been referred to as multisample sphericity. A v-c matrix is spherical if the variances for all possible difference scores are equal. In the repeated measures design a necessary and sufficient condition for a valid test of the within-subjects effect is that $\mathbf{\Sigma}$ is spherical. Results in Box (1954) can be used to show that under violations of sphericity the F test of the within-subjects effect in a repeated measures design is distributed approximately as F with $\varepsilon(K - 1)$ and $\varepsilon(n - 1)(K - 1)$ degrees of freedom. The correction factor ε has a lower bound of $1/(K - 1)$ and an upper bound of one, which is obtained when sphericity holds. Similarly, results in Greenhouse and Geisser (1959) can be used to show that the within-subjects effects in a split plot design are distributed approximately as $F_{\varepsilon(K-1),\ \varepsilon(N-J)(K-1)}$ when the v-c matrices are homogeneous but not spherical.

In the repeated measures design, the t statistic for testing a contrast on the levels of the within-subjects factor is

$$t = \frac{\sum c_k \bar{y}_k}{\sqrt{MS_{B\cdot S} \sum \frac{c_k^2}{n}}}, \tag{1}$$

where $MS_{B\cdot S}$ denotes the error mean square from the analysis of variance. For a split plot design the t statistic for testing a contrast on the elements of $\bar{\boldsymbol{\mu}} = N^{-1}\Sigma n_j \boldsymbol{\mu}_j$ is

$$t = \frac{\sum c_k \bar{y}_{\cdot k}}{\sqrt{MS_{B\cdot S/A} \sum \frac{c_k^2}{N}}}, \tag{2}$$

where $MS_{B\cdot S/A}$ is the error mean square for the within-subjects effects, $\bar{y}_{\cdot k}$ is the *k*th element of $\bar{\mathbf{y}} = N^{-1}\Sigma n_j \bar{\mathbf{y}}_j$, and $\bar{\mathbf{y}}_j$ is the sample mean vector for the *j*th group. Using $MS_{B\cdot S}$ in equation (1) is based on the sphericity assumption in the repeated measures design; using $MS_{B\cdot S/A}$ in equation (2) is based on the multisample sphericity assumption in the split plot design.

Boik (1981) investigated the effects of nonsphericity on the Type I error and power rates of a test using t given in equation (2) and the critical value $\pm t_{\alpha/2,\ (N-J)(K-1)}$. He found that for very small departures from sphericity, the test size can be quite discrepant from α. For $.5 < \varepsilon < 1.0$, increasing the sample size increases the bias of the test. Boik recommended employing the statistic

$$t = \frac{\sum c_k \bar{y}_{\cdot k}}{\sqrt{\frac{\mathbf{c}'\mathbf{S}\mathbf{c}}{N}}}, \tag{3}$$

where **c** is a contrast vector constructed from the c_k and **S** is the pooled v-c matrix and $\pm t_{\alpha/2,\ N-J}$ as the critical value. Boik showed that the power of the test based on equation (2) can be larger or smaller than the power of the test based on equation (3).

The data in repeated measures designs and split plot designs seldom satisfy the sphericity assumption. In addition, the data in a split plot design may have heterogeneous v-c matrices. Several authors have investigated Type I error rates for multiple comparison procedures applied to levels of the within-subjects factor in the repeated measures design and in the split plot design.

In a Monte Carlo study, Maxwell (1980) investigated the impact of violating sphericity on Type I error rates for pairwise comparisons in a repeated measures design. He investigated a procedure in which equation (1) and Tukey's critical value is used and showed that the procedure results in actual Type I error rates above the nominal level, with the discrepancy increasing as ε declines. Maxwell also investigated procedures in which $(\mathbf{c}'\mathbf{S}\mathbf{c}/n)^{1/2}$ replaces the denominator of equation (1), and concluded that this statistic with the Bonferroni critical value results in the best control of the familywise Type I error rate.

Keselman, Keselman, and Shaffer (1991) investigated four procedures that can be used to make multiple pairwise comparisons of the elements of the vector $\bar{\boldsymbol{\mu}}^{*} = J^{-1} \Sigma \boldsymbol{\mu}_j$ in a split plot design. For their test statistic, they replaced the denominator of equation (2) by $[\Sigma_j \ \mathbf{c}'\mathbf{S}_j\mathbf{c}/n_j]^{1/2}$ where $\mathbf{c}$ is the K x 1 vector of contrast weights. They included the following critical values: Studentized maximum modulus, Bonferroni, Studentized range, and an analog of Cochran's

(1964) critical value, each with Satterthwaite degrees of freedom. Keselman, Keselman, and Shaffer concluded that their test statistic with a Studentized range, a Studentized maximum modulus, or a Bonferroni critical value provides reasonable control over the Type I error rates. The Bonferroni critical value resulted in the smallest Type I error rate and the Studentized maximum modulus critical value resulted in a more powerful test than did the Bonferroni critical value. Keselman, Keselman, and Shaffer (1991) found that Type I error rates decreased as ε decreased regardless of the level of heterogeneity.

Keselman (1994) investigated Type I error rates and power for 28 multiple comparison procedures. Keselman adapted Welsch's (1977) step-up procedure by using the test statistic from Keselman, Keselman, and Shaffer (1991) with critical values calculated as described in Welsch. They concluded that Welsch's step-up procedure was robust and tended to be more powerful than the other procedures included in their study.

There has been a limited amount of work on the impact of nonnormality on Type I error rates of procedures for testing multiple comparisons of the levels of the within-subject effect. Keselman, Keselman, and Shaffer found that when the Studentized range critical value was used, there was a tendency for Type I error rate to be larger when the data were sampled from chi-square distributions. There was no consistent trend when the Studentized maximum modulus and Bonferroni critical values were used.

A few authors have investigated the effect of nonnormality on the power of the F-tests by theoretical (David & Johnson, 1951; Srivastava, 1959; Tiku, 1971) and by numerical approaches (Donaldson, 1968; Srivastava, 1959) in the one-way independent samples design. Srivastava (1959) found that skewness has a small effect on the power of the F test but kurtosis leads to a noticeable change in the power particularly in the case of small samples. For example, when the degrees of freedom are 4 and 10 and the population is leptokurtic ($\lambda_3 = 0$, $\lambda_4 = 2.4$), the power of the F test is higher than the normal-theory power up to the point at which power is approximately 0.8. In the region of very high power, the power of the F test under nonnormality is smaller than the normal-theory power. The reverse happens when the kurtosis is negative. Donaldson (1968) investigated the robustness of Type I and II error rates for the F test applied to data sampled from the exponential or lognormal distributions. He concluded that when the test is applied to nonnormal data, it is slightly less powerful than when test is applied to normal data. As the sample size increases, the power differences decline.

There has been a limited amount of work on the effects of nonnormality on the power in the repeated measures design or the split plot design. Rogan, Keselman, and Mendoza (1979) investigated the effect of nonnormality on Type II error rates of the F test in the split plot design using data sampled from chi-square distributions with three degrees of freedom. Rogan, Keselman, and Mendoza found that power rates for the within-subjects main effect and for the within x between interaction were slightly lower for the nonnormal distribution.

Yuen (1975) compared the power of the one-sample Winsorized t, trimmed t and Student's t for long-tailed distributions. The trimmed t and Winsorized t were more powerful than Student's t for sample sizes of 10 or above, when the amount of trimming was appropriate. Yuen and Dixon (1973) investigated the power of two-sample trimmed t test under the assumption of equal population variances. The power of trimmed t never exceeded the power of Student's t under normality, but with sample size of greater than 7 the loss of power was small. For contaminated normal distributions and sample sizes not less than 10, the power of the trimmed t was superior to that of Student's t. Yuen (1974) reported that in the case of unequal sample sizes and variances the power of the trimmed Welch test is superior to the power of Welch's test for long-tailed distributions and it is comparatively more powerful if the larger sample size is associated with the larger variance than with the smaller variance. Wilcox (1992) investigated the effect of nonnormality on Type I error rate and power when comparing two independent groups. Wilcox showed that the power of Welch's procedure drops dramatically as the amount of contamination gets large and Yuen's trimmed Welch test has more power than Welch's test.

Under the violation of the multisample sphericity assumption and/or normal distribution assumption, Wilcox (1993) compared Type I error and power rates of the usual $\tilde{\varepsilon}$-adjusted F test (F) and an $\tilde{\varepsilon}$-adjusted F test using trimmed means and Winsorized variances (D) in repeated measures designs. Wilcox concluded that Type I error rates for D and F were in close agreement. Both procedures were too conservative when n was small and ε was close to one. When ε is relatively small, the F test can have Type I error rate slightly higher than the nominal level when the distributions are normal. In contrast, D has Type I error rate slightly below the nominal level. The power of D is substantially better than F when the distributions have heavy tails and the advantage in using D increases for asymmetric distributions.

The Problem

There has been a limited amount of research on the robustness or power of pairwise multiple comparison procedures when the normality and/or multisample sphericity assumption are violated in a split plot design. However, results in Keselman, Keselman, and Shaffer (1991), Yuen (1974), and Wilcox (1992) suggest a Welch-type test using trimmed means and Winsorized variances may be both robust and powerful.

Purpose of the Study

The purpose of this study is to investigate Type I error rates and power for pairwise multiple comparison procedures when the assumptions of the pairwise comparison procedures are violated in a split plot design with one between-subjects factor.

Significance of the Study

Studies show that long-tailed distributions are common in psychometric and educational measures (Micceri, 1989; Wilcox, 1990). It is known that long-tailed distribution can reduce the power of hypothesis testing procedures based on means (Tukey, 1960; Tukey and McLaughlin, 1963; Yuen, 1974). Further tests based on trimmed means and Winsorized variances maintain power with long-tailed distributions while resulting in a small power loss with normal and short tailed distributions. Consequently, the development and evaluation of multiple comparison procedures using trimmed means and Winsorized variances is of interest in connection with testing contrasts on levels of the within-subjects factor.

CHAPTER 2
LITERATURE REVIEW

Linear Model for the Repeated Measures Design with One Between-Subjects Factor

Consider a design in which there are J levels of a between-subjects factor (groups) with n_j subjects in the jth level. Each subject is measured at each of K levels of a within-subjects factor (treatments or trials). This design is sometimes called a split plot design. Using the traditional notation, the univariate linear model describing this design is

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \pi_{i(j)} + \alpha\beta_{jk} + \beta\pi_{ki(j)} + \varepsilon_{ijk}, \qquad (4)$$

$$i = 1,\ldots,n_j;$$

$$j = 1,\ldots,J;$$

$$k = 1,\ldots,K;$$

where Y_{ijk} is the observed value of subject i within group j at level k of the within-subjects factor, μ is the overall mean, α_j is the effect of group j, β_k is the effect of the level k of the within-subjects factor, $\pi_{i(j)}$ is the effect of subject i nested within group j, $\alpha\beta_{jk}$ is the interaction of group j with level k of the repeated measures factor, $\beta\pi_{jk(i)}$ is the interaction of subject i within group j with level k of the repeated-measures factor and ε_{ijk} is a residual. The observations (Y_{ijk}s) are assumed to be independently and identically distributed within each level j of the between-subjects factor with common mean vector $\boldsymbol{\mu}_j$ and v-c matrix $\boldsymbol{\Sigma}_j$.

In the split plot design, the hypothesis of no within-subjects effect is tested by the ratio

$$F_1 = \frac{MS_B}{MS_{B\cdot S/A}} = \frac{\dfrac{SS_B}{K-1}}{\dfrac{SS_{B\cdot S/A}}{(N-J)\,(K-1)}}, \tag{5}$$

and the hypothesis of no interaction effect is tested by the ratio

$$F_2 = \frac{MS_{AB}}{MS_{B\cdot S/A}} = \frac{\dfrac{SS_{AB}}{(J-1)\,(K-1)}}{\dfrac{SS_{B\cdot S/A}}{(N-J)\,(K-1)}}, \tag{6}$$

where N is the total number of subjects. The ratios are referred to the F distribution with degrees of freedom $(K - 1)$ and $(N - J)(K - 1)$, and $(J - 1)(K - 1)$ and $(N - J)(K - 1)$, respectively.

Validity Conditions for Univariate F Tests

Huynh and Feldt (1970) investigated conditions on the $\boldsymbol{\Sigma}_j$ required for F_1 and F_2 to follow the F distribution described in the preceding paragraph. They showed that necessary and sufficient conditions on the $\boldsymbol{\Sigma}_j$ for the valid F tests of the within-subjects effects are (a) homogeneity of the v-c matrices for a suitable set of orthonormalized variables at all levels of the between-subjects factor, and (b) sphericity of the common v-c matrix. These conditions have been referred to as multisample sphericity. A v-c matrix is spherical if the variances for all possible difference scores are equal. With the homogeneity of the v-c matrices assumption, the variation of the interaction of the subject and the within-subject factors is homogeneous for the groups. This allows pooling over groups to calculate the sum of squares for the interaction of the within-subjects factor and subjects within groups (Brogan & Kutner, 1980). If a design has one within-subjects factor but no between-subjects factors, the condition on $\mathbf{\Sigma}$ that is necessary and sufficient for a valid test of the within-subjects effect is that $\mathbf{\Sigma}$ is spherical (Huynh & Feldt, 1970).

Effects of the Violations of Validity Conditions for Univariate F tests

Robustness against Non-sphericity

Repeated observations seldom satisfy the sphericity assumption (Rogan, Keselman, & Mendoza, 1979). For example, when the within-subjects factor is occasion, adjacent measurements are usually more highly correlated than non-adjacent measurements in split plot and repeated measurements designs (Huynh and Feldt, 1980; Rogan, Keselman, & Mendoza, 1979). Box (1954), studying the repeated measures design, and Imhof (1962), studying the split plot design, found that as the degree of non-sphericity increased, F values for tests involving the within-subjects factor inflated.

Based on theory in Box (1954), it can be shown that when the sphericity assumption is violated in a repeated measures design, the mean square ratio for the within-subjects factor is approximately distributed as F with reduced degrees of freedom. Similarly, theory in Geisser and Greenhouse (1958) can be used to show that mean square ratios for the within-subjects effects are approximately distributed as F with reduced degrees of freedoms. Consequently under violations of sphericity, the within-subjects tests in the repeated measures and split plot designs will be liberal. Several authors have used simulation methods to estimate Type I error rates when sphericity is violated in the repeated measures design (Huynh & Feldt, 1976; Stoloff, 1970) and in the split plot design (Collier, Baker, Mandeville, & Hayes, 1976; Wilson, 1975). These studies have all demonstrated increased Type I error rates when sphericity is violated.

As noted above, theory in Box (1954) can be used to show that under violations of sphericity the F test of the within-subjects main effect is distributed approximately as F with $\varepsilon(K - 1)$ and $\varepsilon(n - 1)(K - 1)$ degrees of freedom. The correction factor ε has a lower bound of $1/(K - 1)$ and an upper bound of one, which is obtained when sphericity holds. The correction factor ε is defined by

$$\varepsilon = \frac{K^2(\overline{\sigma}_{tt}-\overline{\sigma}_{..})^2}{(K-1)(\sum\sum \sigma^2_{ts}-2K\sum \overline{\sigma}^2_{t.}+K^2\overline{\sigma}^2_{..})}, \tag{7}$$

where σ_{ts} ($t \neq s$; $s=1, \ldots, K$) is the *ts*th element of the matrix $\mathbf{\Sigma}$, $\bar{\sigma}_{tt}$ is the mean of the variances, $\bar{\sigma}_{t.}$ is the mean of the *t*th row, and $\bar{\sigma}_{..}$ is the grand mean of all elements in the population v-c matrix.

Two estimates of ε are available. The first, suggested by Greenhouse and Geisser (1959), is

$$\hat{\varepsilon} = \frac{K^2(\overline{S}_{tt}-\overline{S}_{..})^2}{(K-1)(\sum\sum S^2_{ts}-2K\sum \overline{S}^2_{t.}+K^2\overline{S}^2_{..})}. \tag{8}$$

The second suggested by Huynh and Feldt (1976), is

$$\tilde{\varepsilon} = \frac{[N(K-1)\hat{\varepsilon}-2]}{(K-1)[N-J-(K-1)\hat{\varepsilon}]}. \tag{9}$$

Lecoutre (1991) corrected a minor error in the formula for $\tilde{\varepsilon}$:

$$\tilde{\varepsilon} = \frac{(N-J+1)(K-1)\hat{\varepsilon}-2}{(K-1)[N-J-(K-1)\hat{\varepsilon}]}. \tag{10}$$

In a split plot design, the degrees of freedom for the within-subjects effects are approximately $\varepsilon(K - 1)$ and $\varepsilon(N - J)(K - 1)$ when sphericity is violated. In this design $\hat{\varepsilon}$ is calculated by applying equation (5) to the pooled v-c matrix.

A number of studies have used simulated data to estimate Type I error rates for the repeated measures design (Huynh & Feldt, 1976; Stoloff, 1970) and the split plot design (Collier, Baker, Mandeville, & Hayes, 1976; Huynh, 1978; Keselman & Keselman, 1990; Maxwell & Arvey, 1982; Rogan, Keselman, & Mendoza, 1979, Wilson, 1975) when $\hat{\varepsilon}$ and $\tilde{\varepsilon}$ are used to adjust critical values. In general these studies have shown that the performance of the tests depends on ε. When ε is near 1.0, use of the $\hat{\varepsilon}$-adjusted test tends to result in an estimated actual Type I error rate ($\hat{\tau}$) below α and a discrepancy between $\hat{\tau}$ and α that is larger than the discrepancy for the $\tilde{\varepsilon}$-adjusted test. As ε declines, the $\hat{\varepsilon}$-adjusted test achieves better control of τ than the $\tilde{\varepsilon}$-adjusted test. Rough guidelines are that the $\tilde{\varepsilon}$-adjusted test achieves better control than the $\hat{\varepsilon}$-adjusted test for $\varepsilon > .75$, about the same control for $.50 \leq \varepsilon \leq .75$ and worse control for $\varepsilon < .50$. According to Huynh and Feldt, ε is typically above .75 in educational and psychological research.

Robustness against V-C Heteroscedasticity

According to Keselman and Keselman (1990), the majority of studies investigating the robustness of the univariate repeated measurements approach to departures from its assumptions have focused on the effects of non-sphericity of the v-c matrix and not on the effects of heterogeneity of the v-c matrices.

Rogan, Keselman, and Mendoza (1979) investigated the effect of heterogeneous v-c matrices in balanced split plot designs. They found that $\hat{\varepsilon}$- and $\tilde{\varepsilon}$-adjusted univariate procedures are robust with Type I error rates slightly above

α for both within-subjects main effect and between x within interaction.

Keselman and Keselman (1990) investigated the effect of heterogeneous v-c matrices on Type I error rate in balanced and unbalanced split plot designs using three estimates of ε ($\hat{\varepsilon}$, $\tilde{\varepsilon}$, and $\bar{\varepsilon} = \min[1, (\hat{\varepsilon} + \tilde{\varepsilon}) / 2]$, Quintana & Maxwell, 1985) to adjust critical values. They tested the main effect hypothesis $H_{01} : \bar{\mu}_1 = \bar{\mu}_2 = \cdots = \bar{\mu}_K$, where $\bar{\mu}_k$ is the kth element of $\bar{\boldsymbol{\mu}} = N^{-1} \sum n_j \boldsymbol{\mu}_j$, $N = n_1 + n_2 + \cdots + n_J$ and $\boldsymbol{\mu}_j$ is the (K x 1) mean vector for the jth group. They found that heterogeneity affects the size of the test of the between x within interaction when group sizes are unequal, but has little effect on Type I error rate of the within-subjects main effect when group sizes are equal or unequal.

Huynh (1978) extended the $\hat{\varepsilon}$- and $\tilde{\varepsilon}$-adjusted tests to the General Approximation (GA) test and the Improved General Approximation (IGA) test to take the heteroscedasticity of the v-c matrices into account. Huynh conducted a simulation study and compared four approximate tests: the $\hat{\varepsilon}$-adjusted, $\tilde{\varepsilon}$-adjusted, GA, and IGA tests. Huynh included v-c matrices that violated multisample sphericity and both balanced and unbalanced split plot designs. Huynh tested the main effect hypothesis tested in Keselman and Keselman. Huynh found that both the $\hat{\varepsilon}$ approximate and GA procedures are conservative, and the $\tilde{\varepsilon}$ approximate and IGA tests tend to be liberal even though these tests performed reasonably over a range of situations.

Huynh concluded that the IGA tests of the within-subjects main effect and the between x within interaction control the Type I error rate when the v-c matrices are heterogeneous. However, the Type I error rates for the $\tilde{\varepsilon}$-adjusted tests were very similar to those for the IGA tests. Huynh concluded that the $\tilde{\varepsilon}$-adjusted tests are adequate because the IGA test is computationally complex. However, Algina and Oshima (1994) have written a SAS program to calculate the GA and IGA test.

Algina and Oshima (1994) investigated Type I error rates for the $\tilde{\varepsilon}$-adjusted, GA, IGA, and a corrected IGA (CIGA) test based on Lecoutre's (1991) results. They included more severe heterogeneity than did Huynh (1978). Algina and Oshima reported that the $\tilde{\varepsilon}$-adjusted test of the between x within interaction was not robust to heterogeneity and that the IGA and CIGA test provide reasonable control of the Type I error rate. The $\tilde{\varepsilon}$-adjusted, IGA, and CIGA test provided adequate control of the Type I error rate for the test of $H_{01} : \bar{\mu}_1 = \bar{\mu}_2 = \cdots = \bar{\mu}_K$.

Algina and Oshima (in press) investigated the effect of heterogeneous v-c matrices on the Type I error rates of adjusted F tests of the null hypothesis $H_{02} : \bar{\mu}_1^* = \bar{\mu}_2^* = \cdots = \bar{\mu}_K^*$ where $\bar{\mu}_k^*$ is the *k*th element of $\bar{\mu}^* = J^{-1} \Sigma \mu_j$ and μ_j is the mean vector for the *j*th group. They included the $\tilde{\varepsilon}$-adjusted test, the GA test, the IGA test, and the Corrected Improved General Approximation (CIGA) test based on Lecoutre's (1991) result. In contrast to the Keselman and Keselman results in which the $\tilde{\varepsilon}$-adjusted test of H_{01} provided adequate control of τ, the $\tilde{\varepsilon}$-adjusted test of H_{02} provided poor control of the Type I error rate. Algina and Oshima concluded that the CIGA test provides good control over the Type I error rate. The IGA test also provided good control over the Type I error rate, but is slightly more liberal than the CIGA test when both have Type I error rates above $\alpha = .05$ and slightly less conservative when both have Type I error rates below $\alpha = .05$.

Effect of Nonnormality on Type I Error Rates

A number of theoretical (Gayen, 1950; Tiku, 1964) and numerical investigations (Berenson, 1983; Donaldson, 1968; Hack, 1958; Pearson, 1931) of the effect of nonnormality on Type I error rates in the one-way analysis of variance model have been made, but there has been limited amount of work (Algina & Oshima, 1994; Algina & Oshima, in press; Wilcox, 1993) on the impact of nonnormality on type I error rates for the repeated measures design and the split plot design.

Gayen (1950) expressed the distribution of F ratio for nonnormal populations by the first four terms of Edgeworth's series expansion. He showed that Type I error rate is related to the skewness and kurtosis of a distribution. Tiku (1964) expanded probability density functions of the variance components in terms of Laguerre polynomials. He showed that effects due to nonnormality are larger if the skewness is in different directions in different groups. Pearson (1931) studied the effect of nonnormality on Type I error rate of the F test using empirical data with equal group sizes in a one-way independent samples design. Pearson concluded that if the deviation from normality is not extreme, the F test is quite robust. Recently, Berenson (1983) reexamined the issue in a variety of situations when sample sizes are small in a one-way independent samples design. Berenson found that kurtosis has a much greater effect on the Type I error rate of the F test than does skewness. It was concluded that the F test tends to be conservative for the heavy-tailed populations.

Algina and Oshima (1994) investigated the effect of nonnormality on the Type I error rate of the F test using data sampled from multivariate normal and multivariate lognormal distributions in a split plot design. The main effect null hypothesis of the within-subject factor was $H_{01} : \bar{\mu}_1 = \bar{\mu}_2 = \cdots = \bar{\mu}_K$ where $\bar{\boldsymbol{\mu}} = N^{-1} \Sigma n_j \boldsymbol{\mu}_j$ and $\boldsymbol{\mu}_j$ is the mean vector for the jth group. Algina and Oshima included three tests designed for use when multisample sphericity condition is violated: the GA test, the IGA test, and the Corrected Improved General Approximation (CIGA) test. Algina and Oshima showed that, when the sample sizes are equal and v-c matrices are homogeneous, Type I error rates are uniformly lower in the multivariate lognormal conditions than in the multivariate normal conditions for interaction tests. For the main effect tests, the effect of sampling from a multivariate lognormal distribution was that Type I error rates tended to be reduced for ε = 1.00 and to be increased for ε = .40. The effect of sampling from a multivariate lognormal distribution on the within-subjects main effect was complex for the conditions in which v-c matrices were heterogeneous. For K = 4, Type I error rates for ε = 1.00 were reduced and those for ε = .40 were increased. For K = 8 and $\Sigma_1:\Sigma_2:\Sigma_3$ = 1.0:1.5:2.0, Type I error rates for ε = 1.00 and .75 were reduced and those for ε = .40 were increased. For K = 8 and $\Sigma_1:\Sigma_2:\Sigma_3$ = 1:5:9, the Type I error rates for ε = 1.00 were reduced, those for ε = .75 tended to be reduced and those for ε = .40 tended to be increased.

Algina and Oshima (in press) investigated the effect of nonnormality on the Type I error rates for the F test of the main effect of the within-subjects factor in a split plot design under violations of multisample sphericity. Algina and Oshima included four tests: the $\tilde{\varepsilon}$-adjusted test, the GA test, the IGA test, and the CIGA test. The main effect null hypothesis was $H_{02} : \bar{\mu}_1^* = \bar{\mu}_2^* = \cdots = \bar{\mu}_K^*$ where $\bar{\boldsymbol{\mu}}^* = J^{-1} \sum \boldsymbol{\mu}_j$ and $\boldsymbol{\mu}_j$ is the mean vector for the jth group. Algina and Oshima found that the Type I error rate increases as ε decreases for both data sampled from multivariate normal and multivariate lognormal distributions. The effect of nonnormality depended on ε. For $\varepsilon \geq 0.75$, the Type I error rates were lower when sampling from a multivariate lognormal distribution; for ε = .40, Type I error rates were higher.

Effect of Nonnormality on Type II Error Rates

Investigations of the effect of nonnormality on power of the one-way independent samples F test have been conducted by theoretical (David & Johnson, 1951; Srivastava, 1959; Tiku, 1971) and numerical approaches (Donaldson, 1968; Srivastava, 1959).

David and Johnson (1951) obtained approximations of the power function of F test. A numerical investigation showed that the probabilities with the approximation function were generally close to the exact values. Srivastava (1959) expressed the power function of the F test for moderately nonnormal populations by the first four terms of an Edgeworth series. He assumed an equal number of observations in the samples. Srivastava found that skewness had little effect on power of the F test but kurtosis resulted in a noticeable change in the power particularly in the case of small samples. For example, when the degrees of freedom are 4 and 10 and the population is leptokurtic ($\lambda_3 = 0$, $\lambda_4 = 2.4$), the power of the F test is higher than the normal-theory power up to the point at which power is approximately 0.8. In the region of very high power, the power of the F test under nonnormality is smaller than the normal-theory power. The reverse happens when the kurtosis is negative. Tiku (1971) expressed for the power function of the F test with nonnormal data by first six terms of the Laguerre series expansions. A simple two moment approximation for near-normal population distributions was also derived. Tiku found that the effect of moderate nonnormality on the power of F test is unimportant, and the two moment approximation gives accurate approximations in cases of near-normal populations. However, kurtosis had a greater effect on the power than skewness. For the condition of his experiment, $v_1 = 4$ and $v_2 = 20$, the effect of kurtosis was similar to one reported in Srivastava. Donaldson (1968) investigated the robustness of the Type I and II error rates for the F test applied to data sampled from the exponential or lognormal distributions. Donaldson concluded that when the test is applied to nonnormal data, it is slightly less powerful than when test is applied to normal data. As the sample size increases, the power differences decline.

There has been a limited amount of work on the effects of nonnormality on the power in repeated measures design or split plot design. Rogan, Keselman, and Mendoza (1979) investigated the effect of nonnormality on Type II error rates of the F test in the split plot design using data sampled from a chi-square distribution with three degrees of freedom. Rogan, Keselman, and Mendoza found that under nonnormality power rates increase slightly for the within-subjects main effect and for the between x within interaction.

Use of the Trimmed t Test under Nonnormality

The classical statistics such as the sample mean and variance are sensitive to outliers. Trimming and Winsorization refer to the removal and modification of the extreme values of a sample. The purpose of the trimming or Winsorization is to reduce the sensitivity to outliers. Let $y_{(i)}$ denote the *i*th order observation in a random sample of size *N* with $y_{(1)} \leq y_{(2)} \leq \cdot \cdot \cdot \leq y_{(N)}$. The α-trimmed mean for $\alpha = g/N$ or the *g*-times trimmed mean is

$$\bar{y}_{tg} = \frac{1}{(N-2g)} \sum_{i=g+1}^{N-g} y_{(i)} . \qquad (11)$$

The *g*-times Winsorized mean and sum of squared deviations are, respectively,

$$\bar{y}_{wg} = \frac{1}{N} [g y_{(g+1)} + \sum_{i=g+1}^{N-g} y_{(i)} + g y_{(N-g)}] , \qquad (12)$$

and,

$$SSD_{wg} = g(y_{(g+1)} - \bar{y}_{wg})^2 + \sum_{i=g+1}^{N-g} (y_{(i)} - \bar{y}_{wg})^2 + g(y_{(N-g)} - \bar{y}_{wg})^2 . \qquad (13)$$

Tukey and McLaughlin (1963) found that the Winsorized sample variance is a suitable estimate of the variance of the trimmed mean. They suggested the ratio of the trimmed mean to the Winsorized variance

$$t_{tg} = \frac{\bar{y}_{tg} - \mu_0}{[\frac{SSD_{wg}}{(N-2g)(N-2g-1)}]^{1/2}} , \qquad (14)$$

as an alternative to the single sample t test. In equation (14) μ_0 is the hypothesized mean; the null distribution of t_{tg} is approximated by Student's *t* with $N - 2g - 1$ degrees of freedom.

Yuen and Dixon (1973) suggested a trimmed *t* test using

$$t_{tg}=\frac{\bar{y}_{2tg}\ -\ \bar{y}_{1tg}\ -\ (\mu_1-\mu_2)}{[\frac{SSD_{1wg}+SSD_{2wg}}{(N_1-2g_1)+(N_2-2g_2)-2}(\frac{1}{(N_1-2g_x}+\frac{1}{(N_2-2g_2)})]^{1/2}}, \qquad (15)$$

as an alternative to the independent samples t test. In equation (15) $\mu_1 - \mu_2$ is the hypothesized mean difference and numerical subscripts differentiate groups. The distribution of t_{tg} is approximated by Student's *t* with $(N_1 - 2g_1) + (N_2 - 2g_2) - 2$ degrees of freedom.

Yuen (1974) suggested replacing Welch's test on two means by

$$t^* = \frac{(\bar{y}_{1tg}-\bar{y}_{2tg})-(\mu_1-\mu_2)}{[\frac{SSD_{1wg}}{(N_1-2g_1)(N_1-2g_1-1)}+\frac{SSD_{2wg}}{(N_2-2g_2)(N_2-2g_2-1)}]^{1/2}} \qquad (16)$$

which is distributed approximately as Student's t with degrees of freedom *f* calculated from

$$\frac{1}{f} = \frac{c^2}{(N_1-2g_1-1)}+\frac{(1-c)^2}{(N_2-2g_2-1)}, \qquad (17)$$

where

$$c = \frac{\frac{SSD_{1wg}}{(N_1-2g_1)(N_1-2g_1-1)}}{\frac{SSD_{1wg}}{(N_1-2g_1)(N_1-2g_1-1)}+\frac{SSD_{2wg}}{(N_2-2g_2)(N_1-2g_1-1)}}. \qquad (18)$$

Yuen (1975) compared the power of the one-sample Winsorized t, trimmed t and Student's t for long-tailed distributions. The trimmed t and Winsorized t were more powerful than Student's t for sample sizes 10 or above, when the degrees of trimming was appropriate. Yuen and Dixon (1973) investigated the power of the two-sample trimmed t test with equal population variances. The power of trimmed t never exceeded the power of Student's t under normality, but with sample sizes of greater than 7 the loss of power was small. Under contaminated normal distributions and sample sizes not less than 10, the trimmed t was superior to Student's t. Yuen (1974) reported that in the case of unequal sample size and variance the trimmed t is superior to Welch's test for long-tailed distributions and is comparatively more powerful if the larger sample size goes with the larger variance rather than with the smaller variance.

Under the violation of the sphericity assumption and/or normal distribution assumption, Wilcox (1993) compared Type I error and power rates of usual $\tilde{\varepsilon}$-adjusted F test (F) and $\tilde{\varepsilon}$-adjusted F test using trimmed means and Winsorized variances (D) in repeated measures designs. Wilcox included $K = 4$ and 5, $n = 10$, 20, and 40, and 10 and 20 percent trimming. For $K = 5$ the correlation matrices, ρ_{jk}, included $\rho_{jk} = \rho$, $j \neq k$, with $\rho = 0$, 0.1, 0.5, and 0.9, which all have $\varepsilon = 1$. Simulations were also run for all of the correlation matrices considered by Huynh and Feldt (1976), which have $\varepsilon = 0.363$, 0.522, 0.752, 0.831, and 1. For $K = 4$, the same correlations were used but with the last row and column of the correlation matrix removed. Nonnormal distributions were generated using the g-and-h distribution in Hoaglin (1985) and included both symmetric and asymmetric long-tailed distributions. The degrees of freedom for the $\tilde{\varepsilon}$-adjusted F test using trimmed means and Winsorized variances were estimated to be $\nu_1 = \tilde{\varepsilon}(K - 1)$ and $\nu_2 = \tilde{\varepsilon}(K - 1)(n - 2g - 1)$.

Wilcox (1993) concluded that there is close agreement between D and F for Type I error rate. Both procedures were too conservative when n is small and ε is close to one. When ε is relatively small, the F test can have a Type I error rate slightly higher than the nominal level when the distributions are normal. In contrast, D has a Type I error rate slightly below the nominal level. The power of D is substantially better than F when distributions have heavy tails and the advantage in using D seems to increase for asymmetric distributions.

Pairwise Multiple Comparison Procedures

Rejection of the omnibus null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_J$, $H_{01}: \bar{\mu}_1 = \bar{\mu}_2 = \cdots = \bar{\mu}_K$ or $H_{02}: \bar{\mu}_1^* = \bar{\mu}_2^* = \cdots = \bar{\mu}_K^*$ is often followed by pairwise multiple comparisons of the μ_js, to determine which groups are different. There have been several studies of Type I error rates of multiple comparison procedures developed for use in repeated measures or split plot designs.

Maxwell (1980) compared five pairwise multiple comparison procedures that can be used in repeated measures designs.

1. In the Tukey procedure the test statistic is

$$q = \frac{\overline{y}_k - \overline{y}_{k'}}{(MS_{B\cdot S}/n)^{1/2}}, \tag{19}$$

where $\bar{y}_k$ and $\bar{y}_{k'}$ are the sample means for the *k*th and *k*'th level of repeated measures and $MS_{B\cdot S}$ is the error term for the repeated measures ANOVA. The *q* value is compared to the upper α point of the Studentized range distribution based on $(n - 1)(K - 1)$ degrees of freedom. Tukey's procedure is based on three assumptions: (a) normality of the population distribution for each of the *J* populations, (b) equal variances for each of the populations, and (c) equal population variances for all differences between pairs of means, that is, sphericity condition. The validity conditions are the same as those for the omnibus F test in the repeated measures design.

2. In the modified Tukey procedures the test statistic is

$$q = \frac{\overline{y}_k - \overline{y}_{k'}}{\sqrt{(s_k^2 + s_{k'}^2 - 2s_{kk'})/n}}. \tag{20}$$

In one procedure the q value is compared to a critical value of the Studentized Range Statistic with $(n - 1)(K - 1)$; in the other the critical value has $n - 1$ degrees of freedom.

3. In the Bonferroni approach the test statistic is $t = q/\sqrt{2}$ and the critical t has $n - 1$ degrees of freedom and per-comparison level $\alpha_{pc} = 2\alpha_{fw} / K(K - 1)$.

4. In the Simultaneous Confidence Interval (SCI) procedure the test statistic is t and the critical value is

$$\pm\left[\frac{(n-1)(K-1)}{(n-K+1)}F(\alpha, K-1, n-K+1)\right]^{1/2}. \tag{21}$$

The factors in Maxwell (1980) were the number of repeated measures (K = 3, 4, and 5), the sample size (n = 8 and 15), and ε. The quantity ε ranged from 0.49 to 1.00, the variation in ε depended on K. Maxwell concluded that the SCI approach is too conservative, especially when n is small relative to K and that the use of the Tukey procedure under violation of sphericity leads to inflated Type I error rates. Mitzel and Games (1981) reported the same result in a design with n equals 20. Maxwell (1980) recommended the Bonferroni approach as the most powerful among the procedures that control the Type I error rate, $\alpha = .05$.

Keselman and Keselman (1988) investigated four procedures that can be used to make pairwise comparisons of the elements of the vector of $\bar{\boldsymbol{\mu}} = N^{-1} \Sigma n_j \boldsymbol{\mu}_j$ or of $\bar{\boldsymbol{\mu}}^* = J^{-1} \Sigma \boldsymbol{\mu}_j$ in a split plot design. Let $\bar{\mathbf{y}} = N^{-1} \Sigma n_j \bar{\mathbf{y}}_j$ and $\bar{\mathbf{y}}^* = J^{-1} \Sigma \bar{\mathbf{y}}_j$. Keselman and Keselman included the Tukey procedure. To test contrasts on $\bar{\boldsymbol{\mu}}$ the test statistic is

$$q = \frac{\bar{y}_k - \bar{y}_{k'}}{\sqrt{MS_{B \cdot S/A}/N}}. \tag{22}$$

To test contrasts about $\bar{\boldsymbol{\mu}}^*$ the test statistics is

$$q = \frac{\bar{y}_k^* - \bar{y}_{k'}^*}{\sqrt{MS_{B\cdot S/A} \sum_j \frac{1}{n_j}}} . \tag{23}$$

In both cases the critical value is the upper α% point of the Studentized range distribution with $(N - J)(K - 1)$ degrees of freedom. Keselman and Keselman also included a modified Tukey procedure in which the statistic for contrasts on $\bar{\mu}$ is

$$q = \frac{\bar{y}_k - \bar{y}_{k'}}{\sqrt{(s_k^2 + s_{k'}^2 - 2s_{kk'})/N}} , \tag{24}$$

and the test statistic for contrasts on $\bar{\mu}^*$ is

$$q = \frac{\bar{y}_k^* - \bar{y}_{k'}^*}{\sqrt{(s_k^2 + s_{k'}^2 - 2s_{kk'}) (\sum_j \frac{1}{n_j})}} . \tag{25}$$

In equation (24) and equation (25), s_k^2 and $s_{k'}^2$ are pooled variances and $s_{kk'}$ is a pooled covariance. The critical value is the upper α% point of the Studentized range distribution with $N - J$ degrees of freedom. Keselman and Keselman also included two procedures that use $t = q/\sqrt{2}$ as the test statistic. In the first, the Bonferroni procedure, the critical value is a t critical value with $N - J$ degrees of freedom and $\alpha_{pc} = 2\alpha_{fw} / K(K - 1)$. In the second, the SCI procedure, the critical value was

$$\pm \frac{(N-J)(K-1)}{(N-J-K+2)} F(\alpha, K-1, N-J-K+2)^{1/2} . \tag{26}$$

In Keselman and Keselman (1988), J = 3 in all conditions. The population v-c were either equal or $\mathbf{\Sigma}_2 = 3\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_3 = 5\mathbf{\Sigma}_1$.

There were three other factors: (a) $K = 4$ and 8, (b) $\varepsilon = .40$, .75, and 1.0, and (c) n_j = 3, 11, 16; 7, 10, 13; 10, 10, 10; 13, 10, 7; and 16, 11, 3. Keselman and Keselman concluded that when heterogenous v-c matrices are combined with unequal group sizes, none of the four procedures controls the Type I error rates involving comparisons of unweighted means ($\bar{\mathbf{y}}^*$). But, when weighted means ($\bar{\mathbf{y}}$) are used, the Bonferroni and SCI procedure control the Type I error rate. The Bonferroni has more power than the SCI and the power advantage increases as the number of repeated measures levels increases.

Keselman, Keselman, and Shaffer (1991) investigated four procedures that can be used to make multiple pairwise comparisons of the elements of the vector $\bar{\boldsymbol{\mu}}^* = J^{-1} \Sigma \boldsymbol{\mu}_j$ in a split plot design. Keselman, Keselman, and Shaffer included the Studentized maximum modulus critical value, $M_{\alpha,\ c,\ \nu}$, where M is the upper $100(1 - \alpha)$ percentile with parameters $c = K(K - 1)/2$ and Satterthwaite (1941) approximate degrees of freedom ν. Estimated Satterthwaite approximate degrees of freedom are given by

$$\nu = \frac{[\sum_j (s_{jk}^2+s_{jk'}^2-2s_{jkk'})/n_j)]^2}{\sum_j \frac{[(s_{jk}^2+s_{jk'}^2-2s_{jkk'})/n_j]^2}{n_j-1}}. \tag{27}$$

They also included the Bonferroni critical value, $t_{\alpha/(2c),\ \nu}$, where t is the upper $100[1 - \alpha/(2c)]$ percentile of Student's t distribution with degrees of freedom ν given by equation (27), and the Studentized range critical value, $q_{\alpha,\ K,\ \nu}/\sqrt{2}$, with degrees of freedom ν obtained from equation (27). Finally, they included an analog of Cochran's (1964) critical value,

$$C_{\alpha,K} = \frac{\sum_j [q_{\alpha,K,n_j-1}/\sqrt{2}]\cdot[(s_{jk}{}^2+s_{jk'}{}^2-2s_{jkk'})/n_j]}{\sum_j [(s_{jk}{}^2+s_{jk'}{}^2-2s_{jkk'})/n_j]}. \qquad (28)$$

The test statistic in all cases was

$$\frac{\sum_j (\bar{y}_{jk}-\bar{y}_{jk'})}{[\sum_j ((s^2{}_{jk}+s^2{}_{jk'}-2s^2{}_{jkk'})/n_j)]^{1/2}}, \qquad (29)$$

where s_{jk} denotes sample variance for group j and k occasion.

In Keselman, Keselman, and Shaffer (1991) J = 3 in all conditions. Population v-c matrices were equal, or $\boldsymbol{\Sigma}_2 = 1.5\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_3 = 2\boldsymbol{\Sigma}_1$, or $\boldsymbol{\Sigma}_2 = 3\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_3 = 5\boldsymbol{\Sigma}_1$. Two population shapes, normal and chi-square with three degrees of freedom, were included. There were three other factors: (a) $K = 4$ and 8, (b) $\varepsilon = .40, .57, .75$, and 1.0, and (c) n_j = 6, 10, 14; 8, 10, 12; 10, 10, 10; 12, 10, 8; 14, 10, 6 for $N = 30$, and n_j = 9, 15, 21; 12, 15, 18; 15, 15, 15; 18, 15, 12; 21, 15, 9 for $N = 45$. Keselman, Keselman, and Shaffer concluded that the statistic in equation (29) with a Studentized range, a Studentized maximum modulus, or a Bonferroni critical value, based on Satterthwaite's degrees of freedom, provides reasonable control over the Type I error rates. Use of the Bonferroni critical value resulted in the smallest Type I error rate and the Studentized maximum modulus critical value provided a more powerful test than did the Bonferroni critical value.

Keselman, Keselman, and Shaffer (1991) found that the Type I error rate decreased as ε decreased for both normal and nonnormal data regardless of the degree of heteroscedasticity. When the Studentized range critical value was used, Type I error rates were larger for the nonnormal conditions. The tests using the Studentized maximum modulus or the Bonferroni critical values were not consistently affected by nonnormality. With the Studentized range critical value Type I error rates increased as the number of repeated measures levels increased. This was not the case for the tests using the Studentized maximum modulus or the Bonferroni critical value.

Keselman (1994) investigated 28 pairwise multiple comparison procedures:

1. Four adaptations of Hayter's (1986) modification of Fisher's LSD test were used. Four tests of the within-subjects main effect were employed: $\hat{\varepsilon}$-adjusted, $\tilde{\varepsilon}$-adjusted, $\bar{\varepsilon}$-adjusted and Hotelling's T^2. Pairwise comparisons were conducted using equation (29) and $(q_{\alpha,\ K-1,\ \nu})/\sqrt{2}$.

2. An adaptation of Hochberg's (1988) step-up procedure using equation (29).

3. An adaptation of Shaffer's (1986) step-down procedure using equation (29).

4. Four adaptations of Shaffer's sequentially rejective procedure. Four tests of the within-subjects main effect were employed: $\hat{\varepsilon}$-adjusted, $\tilde{\varepsilon}$-adjusted, $\bar{\varepsilon}$-adjusted, and Hotelling's T^2. Pairwise comparisons were conducted using equation (29).

5. Peritz's (1970) procedure using equation (29) as a range statistic and $q/\sqrt{2}$ as critical values, where q denotes a Studentized range critical value.

6. Three adaptations of Peritz's (1970) test using $\hat{\varepsilon}$-adjusted, $\tilde{\varepsilon}$-adjusted, $\bar{\varepsilon}$-adjusted tests and equation (29) for pairwise comparisons.

7. Multiple F tests using $\hat{\varepsilon}$-adjusted, $\tilde{\varepsilon}$-adjusted, $\bar{\varepsilon}$-adjusted tests, with the error mean square computed by using the data for the treatment levels being compared. For pairwise comparison equation (29) was used.

8. Multiple T^2 tests. For pairwise comparison equation (29) was used.

9. An adaptation of the Ryan-Welsch test (Welsch, 1977) using equation (29) and $q/\sqrt{2}$.

10. Four adaptations of Shaffer's (1979) test. Four tests of the within-subjects main effect were employed: $\hat{\varepsilon}$-adjusted, $\tilde{\varepsilon}$-adjusted, $\bar{\varepsilon}$-adjusted, and Hotelling's T^2. Pairwise comparisons were employed using equation (29) and $q/\sqrt{2}$.

11. An adaptation of Welsch's (1977) step-up test. Pairwise comparisons were employed using equation (29) and $q/\sqrt{2}$. Because ν depends on the particular comparison, the

Welsch critical values were forced to follow a monotone sequence. That is, if the critical value for p means was smaller than the critical value for p - 1 means, the former critical value was set equal to the latter.

12. An adaptation of Tukey's test using equation (29) and $q_{\alpha,K,\ \nu}$.

13. A Studentized maximum modulus procedure using equation (29) and $M_{\alpha,\ K(K-1)/2,\ \nu}$.

14. A Bonferroni procedure using equation (29) and $t_{\alpha/K(K-1),\ \nu}$.

In all conditions there were $J = 3$ groups. Let $\mathbf{\Sigma}$ denote the dispersion matrix for the group in which the elements of the dispersion matrix were the smallest. The dispersion matrix was $3\mathbf{\Sigma}$ for one of the other two groups and was $5\mathbf{\Sigma}$ for the remaining group. Keselman manipulated eight factors. Four of the factors were: (a) $K = 4$ or 8, (b) $\varepsilon = .40$ and .75, (c) $N = 30$ and 45, (d) $n_j = 10$, 10, and 10; 8, 10, 12; and 6, 10, 14 for $N = 30$, and 15, 15, and 15; 12, 15, and 18; and 9, 15, and 21 for $N = 45$. Keselman included positive and negative pairings of unequal sample sizes and unequal dispersion matrices and multivariate normal and multivariate nonnormal data. In the nonnormal case canonical transformation of the nonnormal data would result in K uncorrelated variables each with a χ_3^2 distribution. Two configurations of means were included: $\mu_1 = \mu_2 \neq \mu_3 = \mu_4 \neq \cdots \neq \mu_{K-1} = \mu_K$, and $\mu_1 = \mu_2 = \cdots = \mu_K$. Keselman concluded that the Welsch step-up test was robust to nonnormality and tended to have better power than the other multiple comparison procedures.

Dunnett (1982) investigated several pairwise multiple comparison procedures that can be used to compare means from J independent samples. Dunnett considered intervals of the form

$$\tilde{y}_j-\tilde{y}_{j'} \pm A^*_{\alpha,K,\nu_{jj'}}[\frac{\tilde{s}_j^2}{\tilde{n}_j}+\frac{\tilde{s}_{j'}^2}{\tilde{n}_{j'}}]^{1/2}, \qquad (30)$$

where $\tilde{y}_j$, $\tilde{s}_j^2$ denote estimates of location and variance for the *j*th sample, $\tilde{n}_j$ represents effective sample size for the estimator, and $A^*_{\alpha,K,\nu_{jj'}}$ represents a constant which is to be chosen so that the desired joint confidence coefficient $1-\alpha$ is achieved.

Along with the ordinary sample mean and sample variance, five robust estimators of location and the corresponding variance estimate were examined. The robust estimators were wave estimate, bisquare estimate, Hampel estimate, 10% trimmed mean with Winsorized standard deviation, and Tiku's (1967) modified maximum likelihood estimates using 10% trimmed samples. In addition a multiple comparison extension of the Wilcoxon rank sum method was included.

The experiment was conducted for eight symmetric distributions: uniform, normal (0,1), *t* distributions with 10, 4, or 2 degrees of freedom, Cauchy distribution, and two normal outlier distributions. These were .9[N(0,1)] + .1[N(0,3)] and .9[N(0,1)] + .1[N(0,10)]. The experiment had three combinations of K and n: $K = 4$, $n = 10$ or 20 and $K = 8$ and $n = 10$.

Dunnett (1982) concluded that the procedure using the ordinary sample mean and sample variance becomes conservative with data from long-tailed distributions and it should not be used in more than moderate departures from normality. The procedures using trimmed mean and modified maximum likelihood estimate did not become conservative. The nonparametric method should not be used if variance heterogeneity is excessive. The procedure using the wave estimate, bisquare estimate, or Hampel estimate should not be used with data from short-tailed distributions.

Summary and Implications

Based on research by Algina and Oshima (1994), Algina and Oshima (in press), Huynh (1978), Keselman and Keselman (1988), Keselman, Keselman, and Shaffer (1991), and Keselman (1994) the following conclusions can be set forth.

1. Under normality, the $\tilde{\varepsilon}$-adjusted test controls the Type I error rate for the within-subjects main effect on weighted means even when multisample sphericity is violated. There does not appear to be any advantage to using the GA or IGA tests. Although the available results are limited, the

conclusion appears to be true when the data come from a multivariate lognormal distribution.

2. When the unweighted means are used the $\tilde{\varepsilon}$-adjusted test does not exhibit good control of the Type I error rate of the within-subjects main effect for unequal samples sizes and heterogeneous v-c matrices. However, when sample sizes are equal the test provides good control of the Type I error rate even when the v-c matrices are heterogeneous. When the sample sizes are not equal the IGA and CIGA tests control the Type I error rate much better than does the $\tilde{\varepsilon}$-adjusted test.

3. For the multiple comparison tests with the weighted means, Type I error rates of Tukey and modified Tukey (Maxwell, 1980) pairwise multiple comparison procedures are affected under the violated multisample sphericity, but the Bonferroni and SCI procedures have a good control of the Type I error rate. The Bonferroni procedure has more power than the SCI procedure does and the power advantage increases as the number of repeated measures levels increases.

4. When heterogeneous v-c matrices are combined with unequal group sizes, none of the Tukey, modified Tukey (Maxwell, 1980), Bonferroni, or SCI has a good control over the Type I error rates involving comparisons of unweighted means. Type I error rate decreases as ε decreases for both normal and nonnormal data regardless of the degree of heteroscedasticity. There is a variety of multiple comparison procedures that control the Type I error rate for comparisons

of unweighted means. Keselman (1994) reports that an adaptation of the Welsch (1977) step-up test tends to have the best power.

Studies by Dunnett (1980), Tukey and McLaughlin (1963), Wilcox (1992, 1993), Yuen (1974), and Yuen and Dixon (1973) indicate that normal theory tests become conservative and have reduced power with data from long-tailed distributions. Replacing means by trimmed means and variance by Winsorized variances tends to solve these problems.

The purpose of the current study is to investigate Type I error rates and power of multiple comparison procedures that can be used in split plot designs. Both normal-theory procedures and alternatives constructed by replacing means by trimmed means and variances by Winsorized variances are considered.

CHAPTER 3
METHODOLOGY

In this chapter, the multiple comparison procedures and their test statistics, the design, and the simulation procedure are described. The design is based upon a review of relevant literature and upon the consideration that the experimental conditions used in the simulation should be similar to those found in educational research.

Multiple Comparison Procedures

Three types of multiple comparison procedures were employed in the study: (a) the Bonferroni procedure, (b) Shaffer's modified sequentially rejective procedure, and (c) Welsch's step-up procedure. The implementation of the procedures is described in the next several paragraphs.

Bonferroni Procedures

Three variations of the Bonferroni procedure were investigated. In the first, the test statistic was

$$\frac{\sum_{j} [\bar{y}_{jk}-\bar{y}_{jk'}]/J}{\sqrt{\sum_{j} [(s_{jk}{}^2+s_{jk'}{}^2-2s_{jkk'})/n_j]/J^2}}. \qquad (31)$$

In equation (31) $\bar{y}_{jk}$ and s^2_{jk} are the mean and variance of y_k at the jth level of the between-subjects factor; $s_{jkk'}$ is the covariance between y_k and $y_{k'}$ at the jth level of the between-subjects factor. The critical value was $\pm t_{\alpha_{pc}/2,\nu_1}$, where $\alpha_{pc} = 2\alpha_{fw} / [K(K-1)]$ and

$$\nu_1 = \frac{[\sum_j (s_{jk}{}^2+s_{jk'}{}^2-2s_{jkk'})/n_j)]^2}{\sum_j \frac{[(s_{jk}{}^2+s_{jk'}{}^2-2s_{jkk'})/n_j}{n_j-1}}. \tag{32}$$

Subsequently, this procedure is referred to as the Bonferroni (B) procedure.

In the second procedure $\bar{y}_{jk}$ and $\bar{y}_{jk'}$ were replaced by g-times trimmed means and s^2_{jk}, $s^2_{jk'}$, and $s_{jkk'}$ were replaced by Winsorized variances and covariances, respectively. The Winsorized covariance is defined as

$$s_{jkk'}{}^* = \sum_i \frac{(y_{ijk}{}^*-\bar{y}_{jk}{}^*)(y_{ijk'}{}^*-\bar{y}_{jk'}{}^*)}{(n_j-2g-1)}, \tag{33}$$

where y^*_{ijk} is a Winsorized data point and $\bar{y}^*_{jk}$, $\bar{y}^*_{jk'}$ are Winsorized means. The critical value was $\pm t_{\alpha_{pc}/2,\nu_2}$. The degrees of freedom ν_2 are obtained by replacing s^2_{jk} and $s_{jkk'}$ by Winsorized variances and covariances, respectively; the quantities $n_j - 1$ were replaced by $n_j - 2g - 1$. The second procedure is referred to as the raw score trimmed Bonferroni procedure (B_{RT}).

Let $d_{kk'} = y_k - y_{k'}$. In the third procedure the test statistic was

$$\frac{\sum_j \bar{d}^*_{jkk'}/J}{\sqrt{\sum_j [s^{*2}_{jkk'}/n_j]/J^2}}, \tag{34}$$

where $\bar{d}^*_{jkk'}$ is the trimmed mean of $d_{jkk'} = y_{jk} - y_{jk'}$ and $s^*_{jkk'}$ is the Winsorized standard deviation of $d_{jkk'} = y_{jk} - y_{jk'}$. The critical value was $\pm t_{\alpha_{pc}, \nu_3}$ where

$$\nu_3 = \frac{[\sum_j s^{*2}_{jkk'}/(n_j-2g)]^2}{\sum_j \frac{[s^{*2}_{jkk'}/(n_j-2g)]^2}{n_j-2g-1}}. \tag{35}$$

The third procedure is called the trimmed difference score Bonferroni procedure (B_{DT}).

The Bonferroni procedures were included primarily because multiple comparison procedures using the Bonferroni critical value are well known and commonly employed.

Modified Sequentially Rejective Bonferroni procedures

Six variations on Shaffer's modified sequentially rejective Bonferroni (MSRB) procedure were included in the study. The steps in the MSRB procedure will be described in the context of the first variation included in the study. Then the other MSRB procedures will be described.

The MSRB procedure is a modification of the Bonferroni procedure which is implemented in several steps. The first step is to test $H_{02} : \bar{\mu}_1^* = \cdots = \bar{\mu}_K^*$. In the first MSRB procedure, H_{02} was tested by using the $\tilde{\varepsilon}$-adjusted test. If H_{02} is not rejected, all pairwise comparisons are declared nonsignificant. If H_{02} is rejected, test statistics for the $K(K - 1)/2$ pairwise comparisons are calculated and ranked by their p values from smallest to largest. In the first MSRB procedure the test statistic was calculated by using equation (31) and its degrees of freedom were calculated by using equation (32). The critical value for the test statistic with rank i ($i = 1, \cdot \cdot \cdot, K(K - 1)/2$) is $\pm t_{\alpha_{fw}/2c_i, v_1}$. In this study, K = 4; for $K = 4$, $K(K - 1)/2 = 6$ and there are six c_i: $c_1 = c_2 = c_3 = c_4 = 3$; $c_5 = 2$; $c_6 = 1$. Testing is conducted in steps. If the test statistic in the *i*th step is not significant, all subsequent steps are declared nonsignificant. The first MSRB procedure is denoted by $S(\tilde{\varepsilon})$.

The potential advantage of $S(\tilde{\varepsilon})$ over B is that the critical values for the $S(\tilde{\varepsilon})$ procedure are uniformly less extreme than are the critical values for B. The potential disadvantage is that a lower ranked test statistics can be declared nonsignificant because of a nonsignificant test in a prior step. Nevertheless, MSRB procedures tend to be more powerful than Bonferroni procedures (Keselman, 1994).

In the second MSRB procedure, H_{02} is tested by using a trimmed $\tilde{\varepsilon}$-adjusted test. That is, in the calculation of F and $\tilde{\varepsilon}$, trimmed means replace means and Winsorized variances and covariances replace variances and covariances. In subsequent steps the test statistics and degrees of freedom employed in B_{RT} are used. The second MSRB procedure is denoted by $S(\tilde{\varepsilon})_{RT}$. The third MSRB procedures, denoted by $S(\tilde{\varepsilon})_{DT}$, uses the trimmed

$\tilde{\varepsilon}$-adjusted test in the first step; in subsequent steps the test statistics and degrees of freedom used in B_{DT} are used.

The fourth MSRB procedure is the same as the first, except the CIGA test developed by Algina and Oshima (in press) replaces the $\tilde{\varepsilon}$-adjusted test. The fourth MSRB test is denoted by S(C). In the CIGA test the statistic for testing H_0^* is the usual F statistic given by equation (5). The critical value is $\hat{b}F_{\alpha,\tilde{h}',\tilde{h}}$ where $\hat{b}$, $\tilde{h}'$, and $\tilde{h}$ are described as follows.

Let **R** denote a [(*K* - 1) x *K*] matrix with rows that are orthonormal contrast vectors, $\bar{\mathbf{x}}_j$ the (*K* x 1) sample mean vector for group *j*, $\mathbf{S}_j$ the (*K* x *K*) sample dispersion matrix for group *j*. In order to adjust critical value for violations of multisample sphericity, the GA test uses $\hat{b}F_{\alpha,\tilde{h}',\tilde{h}}$ as the critical value. The arguments of the critical value are defined in terms of the $\mathbf{S}_g$ matrices. Define

$$\boldsymbol{S} = \frac{\sum fxmj\boldsymbol{S}_j/n_j}{\sum_j 1/n_j}. \tag{36}$$

Based on Box (1954) and using the results in Huynh (1978) Algina and Oshima (in press) showed that for the test of H_{02} : $\bar{\mu}_1^* = \cdots = \bar{\mu}_K^*$

$$\hat{b} = \frac{(N-J)\,tr(\boldsymbol{RSR}')}{\sum_j (n_j-1)\,tr(\boldsymbol{RS}_j\boldsymbol{R}')}, \tag{37}$$

$$\hat{h}' = \frac{[tr(\boldsymbol{RSR}')]^2}{tr(\boldsymbol{RSR}')^2}, \tag{38}$$

and

$$\hat{h} = \frac{[\sum_j (n_j-1)\, tr(\boldsymbol{RS}_j\boldsymbol{R}')]^2}{\sum_j (n_j-1)\, tr(\boldsymbol{RS}_j\boldsymbol{R}')^2}. \tag{39}$$

Huynh proposed the IGA test which replaces $\hat{h}$ and $\hat{h}'$ by statistics that are intended to reduce the underestimation which occurs when multisample sphericity nearly holds. Let $A_j = tr(\mathbf{RS}_j\mathbf{R}')$ and $B_j = tr(\mathbf{RS}_j\mathbf{R}')^2$. In the IGA test, $\hat{h}$ is replaced by $\tilde{h} = \hat{\eta}/\hat{\delta}$ where

$$\hat{\eta} = \sum_j \frac{(n_j-1)^3}{(n_j+1)(n_j-2)}(n_jA_j^2-2B_j) + \sum\sum_{j\neq j'} (n_j-1)(n_{j'}-1)A_jA_{j'}, \tag{40}$$

and

$$\hat{\delta} = \sum_j \frac{(n_j-1)^2}{(n_j+1)(n_j-2)}[(n_j-1)B_j-A_j^2]. \tag{41}$$

Based on the expression for $\tilde{\varepsilon}$ (Huynh and Feldt, 1976) Huynh suggested replacing $\hat{h}'$ by

$$\tilde{h}' = \frac{N\hat{h}' - 2}{N - J - \hat{h}'}. \tag{42}$$

Lecoutre (1991) presented a correction to the IGA test by replacing N in the numerator of $\tilde{h}'$ by $N - J + 1$. It is called the CIGA test.

In the fifth MSRB test, denoted by $S(C)_{RT}$, a raw score trimmed CIGA test is used to test H_{02}. In this test, the means, variances and covariances used in the CIGA test were replaced by trimmed means and Winsorized variances and covariances, respectively. The test statistics and degrees of freedom employed in the subsequent steps are the same as in $S(\tilde{\varepsilon})_{RT}$. In the sixth MSRB procedure, denoted by $S(C)_{DT}$, the trimmed CIGA test is again used in the first step. The test statistics and degrees of freedom employed in the subsequent steps are the same as in $S(\tilde{\varepsilon})_{DT}$.

Welsch's Step-up Procedures

Two variations of Welsch's step-up procedure were included in the study. Steps in Welsch's step-up procedure will be described in the context of the first variation included in the study. Then the other procedures will be described.

In Welsch's step-up procedure the means for all *K* levels of the within-subjects factor are ordered from smallest to largest. The procedure begins by testing all 2-ranges, $(\bar{Y}^*_{(k+1)} - \bar{Y}^*_{(k)})$ $(1 \le k \le K-1)$, of the ordered treatment means of each level of the within-subjects factor, $\bar{Y}^*_{(1)} \le \bar{Y}^*_{(2)} \le \cdot \cdot \cdot \le \bar{Y}^*_{(K)}$. If the 2-range is significant, the members of the corresponding pair of treatments are declared different. Also declare the *p*-ranges containing that 2-range significant and all sets of treatments which contain that significant subset as heterogeneous by implication without further tests. If at least one 2-range is not significant, then proceed to test 3-ranges. In general, test a *p*-range $(\bar{Y}^*_{(p+k-1)} - \bar{Y}^*_{(k)})$ $(1 \le k$

$\le K - p + 1$, $2 \le p \le K$), if that p-range is not declared significant by implication at earlier steps. If $\bar{Y}^*_{(p+k-1)} - \bar{Y}^*_{(k)}$ is significant, then declare the pair of treatments corresponding to $\bar{Y}^*_{(p+k-1)}$ and $\bar{Y}^*_{(k)}$ as different. Also declare all sets of treatments containing that subset of p treatments as heterogeneous and the corresponding q-ranges containing that p-range for all $q > p$ as significant by implication without further tests. Welsch proposed using $\alpha_p = p\alpha/K$ for $2 \le p \le K - 2$, and $\alpha_{K-1} = \alpha_K = \alpha$. Welsch (1977b) tabulated the critical values $\xi_{\alpha_p,\nu}$ ($2 \le p \le K$) based on Studentized range statistics for k = 2, 3, · · ·, 10, ν = 5, 6, ···, 20, 24, 30, 40, 60, 120, ∞, and $\alpha = .05$ which he obtained using Monte Carlo simulations.

In the first variation of Welsch's step-up procedure, denoted by W, the test statistic was calculated using equation (31). The critical value is $\xi_{\alpha_p,\nu_1}/\sqrt{2}$. In the second variation of Welsch's procedure the ordered means used in the first variation are replaced by $\Sigma_j\ \bar{Y}_{tjk}/J$ where $\bar{Y}_{tjk}$ is the trimmed mean for the jkth cell of the design. The test statistic for comparing means is the test statistic used in B_{RT}. The critical value is $\xi_{\alpha_p,\nu_2}/\sqrt{2}$. This procedure is referred to as the raw score trimmed Welsch procedure (W_{RT}). In both procedures the critical values were forced to follow a monotone sequence. That is ξ_{α_p,ν_s} was set equal to ξ_{α_{p-1},ν_s} whenever $\xi_{\alpha_p,\nu_s} < \xi_{\alpha_{p-1},\nu_s}$ ($3 \le p \le K$ and s = 1, 2, 3).

The potential advantage of Welsch's step-up procedures is better power than Shaffer's procedures because the step-up procedures declare significance by implication; in contrast Shaffer's procedures declare nonsignificance by implication.

Design

Six factors were manipulated to investigate the performance of the various multiple comparison procedures. These factors were (a) distribution of the generated data, (b) sphericity of the common v-c matrix, (c) degree of the heteroscedasticity of the v-c matrices, (d) total sample size, (e) degree of the sample size ratio, and (f) the configuration of the nonnull means. There were $J = 2$ levels for the between-subjects factor and $K = 4$ levels for the within-subjects factor in all conditions.

Distribution of the Generated Data

In one-third of the conditions the data in each group were sampled from multivariate normal distribution. The g-and-h method (Hoaglin, 1985) was used to simulate nonnormal data. Micceri (1989) reported distributional characteristics for 440 data sets from education and psychological research. He reported that 32.5% had at least one heavier tail than the extreme contamination and 16.6% had heavier tail than double exponential distribution (kurtosis = 6.0). The extreme contamination represents more than 100 times the expected observations of normal distribution over 3 standard deviations from the mean. Based on Micceri, in one-third of the conditions the data in each group were sampled from the g-and-h distribution with $g = 0$ and $h = .109$ approximating double exponential distribution. The results in Wilcox (1993) indicate that methods based on trimmed means are more powerful than those based on untrimmed means when the data are very long tailed. Based on Wilcox's results, in one-third of the conditions the data in each group were drawn from g-and-h distribution with $g = 0$ and $h = .35$.

Sphericity of the Common V-C Matrix (ε)

Box's (1954) correction factor, ε, was used to quantify the degree of departure from the sphericity assumption. Two levels of ε were included to investigate the effects of nonsphericity. These were $\varepsilon = .96$, and $\varepsilon = .75$. Dispersion matrices were taken from Keselman and Keselman (1990). Table 1 contains the dispersion matrices for each value of ε. To achieve comparability across the simulation conditions, dispersion matrices with average diagonal value of 10 and average off-diagonal value of 5 were used.

Degree of Heteroscedasticity of the V-C Matrices

Conditions in which the v-c matrices were equal in both groups and conditions in which the v-c matrices were not equal were included. In the former conditions, the v-c matrix for both groups was one of the matrices in Table 1. Two levels of dispersion matrix inequality were included. In the first, $\mathbf{\Sigma}_1 = (2/3)\mathbf{\Sigma}$ and $\mathbf{\Sigma}_2 = (4/3)\mathbf{\Sigma}$ where $\mathbf{\Sigma}$ is one of the matrices in

Table 1

V-C Matrices for ε = .96 and .75

ε	v-c matrix			
0.96	12.0	6.0	5.0	5.0
		10.0	5.0	4.0
			10.0	5.0
				8.0
0.75	18.0	8.0	6.0	4.0
		8.0	5.0	4.0
			7.0	3.0
				7.0

Table 1. In the second, $\mathbf{\Sigma}_1 = (1/3)\mathbf{\Sigma}$ and $\mathbf{\Sigma}_2 = (5/3)\mathbf{\Sigma}$. These levels of dispersion are referred to as 1:1, 1:2 and 1:5 in subsequent tables.

Total Sample Size (N)

Two levels of total sample sizes were investigated. These were $N = 40$ and $N = 60$.

Sample Size Ratio (n_1, n_2)

Conditions with equal sample sizes for each between-subjects level and conditions with unequal sample sizes were investigated. Both moderate and large degrees of sample size inequality were included. For $N = 40$ the sample size combinations were n_1, n_2 = 28, 12; 20, 20; and 12, 28. For $N = 60$ the sample size combinations were n_1, n_2 = 42, 18; 30, 30; and 18, 42. Thus, there were three levels of the sample size combinations factor for each total sample size $N = 40$ and $N = 60$, which are referred to as levels A to C in subsequent tables. For those conditions involving unequal between-subjects group sizes and unequal v-c matrices, both positive and negative pairings of between-subjects sample sizes and v-c matrices were included. A positive pairing refers to the case in which the group with larger sample size, n_j, is associated with the v-c matrix containing the larger element values; a negative pairing refers to the case in which the larger sample size, n_j, is associated with the v-c matrix with the smaller element values.

The Configuration of the Nonnull Means

To investigate power, three configurations of nonnull means were examined (Ramsey, 1978): (a) a minimum range configuration, which occurs when half of the means are equal and the remaining half are also equal but different from the first half; (b) a maximum range configuration, which occurs when there is one mean at each extreme of the range and the remaining means equal the average of these two extreme values; and (c) an equally spaced range configuration, which occurs when the means are equally spaced across the range.

When $\mathbf{\Sigma}$ satisfies $\mathbf{C'\Sigma C} = \lambda \mathbf{I}_{K-1}$ with $\mathbf{C'C} = \mathbf{I}_{K-1}$, the noncentrality parameter in the split plot design with two levels in between-subjects factor can be shown to be

$$\delta^2 = \frac{n_1 n_2}{n_1+n_2} \frac{\sum_{k=1}^{K} (\mu_k^{*2} - \bar{\mu}^2)}{\lambda}, \tag{43}$$

where $\mu_k^* = (\mu_{1k} + \mu_{2k})/2$, μ_{ik} is the mean of the *i*th level of the between-subjects factor and *k*th level of the within-subject factor, n_i is the number of sample for the *i*th level of the between-subjects factor, and $\bar{\mu}$ is the mean of μ_k^*s. For each sample size combination, the value of the noncentrality parameter necessary to produce omnibus power of .50 for the case in which $\alpha = 0.05$ and v-c matrices are equal was calculated. Then, the means for each level of the within-subjects factor were determined by using

$$\sum_{k=1}^{K} (\mu_{k}^{*2} - \overline{\mu}^{2}) = \frac{n_1 + n_2}{n_1 n_2} \lambda \delta^{2}. \tag{44}$$

The μ_k^*s are assumed to sum to zero and monotonically increase from k = 1 to K. The resulting means are presented in Table 2.

Simulation Procedure

Data Generation

The data for each condition that involved multivariate normal data were generated using the following steps:

1. For the *j*th level of the between-subjects factor $\mathbf{Z}_j$, an n_j x 4 matrix of independent normally distributed variates was generated. The NORMAL function in SAS was used to generate all variates.

2. The matrix $\mathbf{Z}_j$ was transformed to

$$\mathbf{X}_j = \mu + d_j \mathbf{Z}_j \mathbf{U}', \tag{45}$$

where $\mathbf{U}$ is a lower triangular matrix satisfying the equality $\mathbf{\Sigma}_1 = \mathbf{UU}'$, d_j is a constant selected to simulate the required degree of heteroscedasticity, and μ is an n_j x 4 matrix of means selected to simulate the required configuration of means.

The data from nonnormal distributions were generated using the g-and-h distribution suggested by Tukey (1977) and developed by Hoaglin (1985). This family of distributions is attractive in simulation studies because those distribution

Table 2

Means of Repeated Measures for each Sample Size Arrangement (S) to get 50 percent Power in Omnibus F Test when all Assumptions for omnibus F Test are satisfied

configuration of the nonnull means	S	μ_1	μ_2	μ_3	μ_4
minimum range means	a,c	-0.470918	-0.470918	0.470918	0.470918
	b	-0.431603	-0.431603	0.431603	0.431603
	d,f	-0.382238	-0.382238	0.382238	0.382238
	e	-0.350327	-0.350327	0.350327	0.350327
maximum range means	a,c	-0.665978	0	0	0.665978
	b	-0.610379	0	0	0.610379
	d,f	-0.540567	0	0	0.540567
	e	-0.495438	0	0	0.495438
equal space means	a,c	-0.631802	-0.210601	0.210601	0.631802
	b	-0.579056	-0.193019	0.193019	0.579056
	d,f	-0.512827	-0.170942	0.170942	0.512827
	e	-0.470013	-0.156671	0.156671	0.470013

Note. a:n_j = 12 and 28; b:n_j = 20 and 20; c:n_j = 28 and 12; d:n_j = 18 and 42; e:n_j = 30 and 30; f:n_j = 42 and 18.

shapes are determined by a small number of parameters, a wide spectrum of distributions can be approximated, and simulated observations can be easily generated from independent and identically distributed normal deviates.

The g-and-h distributions are generated by a single transformation of the standard normal variable Z_{ij},

$$X_{ij} = \frac{\exp(gZ_{ij})-1}{g}\exp(hZ_{ij}^2/2), \tag{46}$$

where Z_{ij} is a standard normal variable, g is a real constant, and h is a nonnegative real constant. The g-and-h distributions are defined in terms of their quantile function relative to the standard normal distribution so that if Z_{ij} in equation (46) has the value of the pth quantile of the standard normal distribution, then X_{ij} is the pth quantile of a g-and-h distribution with the specified values of g and h.

Martinez and Iglewicz (1984) and MacGillivray (1992) investigated some properties of this family. The g-and-h distribution for the subfamily with $g = 0$ is

$$X_{ij} = Z_{ij}\exp(hZ_{ij}^2/2), \tag{47}$$

because

$$\frac{\exp(gZ_{ij})-1}{g} = Z_{ij} + gZ_{ij}^2/2! + g^2Z_{ij}^3/3! + \cdots. \tag{48}$$

For the case $g = h = 0$, it follows easily that $X_{ij} = Z_{ij}$. The case $g = 0$ corresponds to a symmetric distribution, and increased g results in greater skewness to the right for $g > 0$; the same amount of skewness to the left can be obtained by choosing $g < 0$ with same absolute value. The parameter h determines heavy tailedness. As h increases, heavy tailedness increases as well. For the case $h = 0$ the g-and-h distribution,

$$X_{ij} = \frac{\exp(gZ_{ij})-1}{g}, \tag{49}$$

coincides with the lognormal distribution. In addition to the g-and-h distributions allowing only for longer-tailed distributions than the normal distribution, a possible drawback of the g-and-h distributions is the dependence of the spread function on g. The skewness depends only on g and increases with the absolute value of g, but kurtosis increases with h and with the absolute value of g. Table 3 cited from Martinez and Iglewicz (1984) shows the values of g and h that approximate a selected set of well known distributions in comparison with the usual coefficients of skewness and kurtosis.

The nonnormal data were generated using the g-and-h distribution as the following steps:

1. For the jth level of the between-subjects factor $\mathbf{Z}_j$, an n_j x 4 matrix of independent normally distributed variates was generated. The NORMAL function in SAS was used to generate all variates.

2. An n_j x 4 matrix $\mathbf{X}_j^*$ was constructed by applying,

$$X_{ij}^* = Z_{ij}\exp(hZ_{ij}^2/2). \tag{50}$$

Table 3

Coefficients of skewness and kurtosis and corresponding values of g and h for some selected distributions

Distribution	skewness	kurtosis	g	h
Uniform	.00	1.80	.000	-.244
Normal	.00	3.00	.000	.000
t_{10}	.00	4.00	.000	.058
Logistic	.00	4.20	.000	.065
Double Exponential	.00	6.00	.000	.109
Cauchy	-	-	.000	.970
Beta (2,3)	.29	2.36	.142	-.101
Weibull; k = 2	.63	3.25	.243	-.042
χ^2_{10}	.89	4.20	.303	-.017
χ^2_4	1.41	6.00	.502	-.046
Exponential	2.00	9.00	.760	-.098
Lognormal	6.18	113.94	1.000	.000

3. The n_j x 4 matrix $\mathbf{X}_j^*$ was transformed to

$$\mathbf{X}_j^* = \mu + d_j\mathbf{X}_j\mathbf{U}', \tag{51}$$

where μ, d_j, and $\mathbf{U}'$ are defined as in the second step of the procedure for generating multivariate normal data.

Evaluating the Performance of Test Procedures

Type I error rates were obtained under conditions where the population mean vector, μ, was the null vector. The Type I error was defined the probability that at least one of the pairwise comparisons was statistically significant. Three different types of power were computed: (a) the average per-pair power rate, where per-pair power is the probability of detecting a non-zero pairwise difference for a particular pairwise comparison; (b) the all-pairs power rate, that is, the probability of detecting all non-zero pairwise differences; and (c) the any-pairs power rate, that is, the probability of detecting at least one non-zero pairwise difference. For each condition 5000 replications were performed. The proportions of the replications that resulted in a significant test at $\alpha = .01$, $\alpha = .05$, and $\alpha = .10$ were reported. Welsch's (1977) step-up procedure was conducted only at $\alpha = .05$.

Summary

Three distribution types (g = 0 and h = 0, g = 0 and h = .109, and g = 0 and h = .35), two levels of sphericity of the common v-c matrix (ε = .96 and .74), three levels of the degree of the heterogeneity of the v-c matrices ($\Sigma_1:\Sigma_2$ = 1:1, 1:2, and 1:5), two levels of total sample size (N = 40, and 60), three levels of sample size ratio (n_1, n_2) = 28, 12; 20, 20; and 12, 28 for N = 40, and 42, 18; 30, 30; and 18, 42 for N = 60), and a configuration of null means and three configuration of nonnull means combine to give 432 experimental conditions.

Three variations of the Bonferroni procedures were included: (a) the Bonferroni (B), (b) the raw score trimmed Bonferroni (B_{RT}), and (c) the trimmed difference score Bonferroni (B_{DT}). Six variations of the Shaffer's modified sequentially rejective Bonferroni (MSRB) procedures were included: (a) Shaffer's MSRB procedure using $\tilde{\varepsilon}$-adjusted test for omnibus test ($S(\tilde{\varepsilon})$), (b) Shaffer's MSRB procedure using $\tilde{\varepsilon}$-adjusted test for omnibus test with the raw score trimmed data ($S(\tilde{\varepsilon})_{RT}$), (c) Shaffer's MSRB procedure using $\tilde{\varepsilon}$-adjusted test for omnibus test with the trimmed difference score ($S(\tilde{\varepsilon})_{DT}$), (d) Shaffer's MSRB procedure using the trimmed CIGA test for omnibus test (S(C)), (e) Shaffer's raw score trimmed MSRB procedure using the trimmed CIGA test for omnibus test ($S(C)_{RT}$), and (f) Shaffer's trimmed difference score MSRB procedure using the trimmed CIGA test for omnibus test ($S(C)_{DT}$). Two variations of Welsch's step-up procedures were included: (a) Welsch's step-up procedure (W) and (b) raw score trimmed Welsch procedure (W_{RT}).

CHAPTER 4
RESULTS AND DISCUSSION

In this chapter analyses of Type I error rates and power are presented for tests conducted at $\alpha = .05$. The pattern of results for tests conducted at $\alpha = .01$ and for $\alpha = .10$ was similar. Results are reported in five sections: (a) Type I error rates for a completely true null hypothesis, (b) Type I error rates for a partially true null hypothesis, (c) any-pair power, (d) average per-pair power, and (e) all-pairs power. Each subsection begins with a table of percentiles of the distributions of Type I error rates or power. Then a summary of the analysis of variance (ANOVA) used to investigate the effect of the within-subjects factor and the between-subjects factors on Type I error rates or power is presented. The within-subjects factor was test (T) and the between-subjects factors were distribution of the generated data (D), sphericity of the common v-c matrix (ε), degree of v-c matrices heteroscedasticity ($\Sigma_1:\Sigma_2$), total sample size (N), sample size arrangement (n_1/n_2), and configuration of the means (CM).

Because many of the factors that affect Type I error rates and power were included in the study, each ANOVA yielded a substantial number of significant effects. To compare the size of the effects $\hat{\omega}^2$ (Myers, 1979) was calculated for each effect. A mean square component was computed for each effect in each ANOVA. The mean square component, $\hat{\theta}_i^2$, for an effect was computed using the formula [(MSF$_i$ - MSE) / x], where MSF$_i$ was the mean square for the given effect, MSE was the between-subjects mean square error when the given effect was between-subjects effects and the within-subjects mean square error when the given effect was the within-subjects effect or the interaction of the within-subjects factor and between-subjects factor(s), and x was the number of levels for the factors not included in the given effect. Negative mean square components were set to zero. For each effect with a non-negative mean square component, $\hat{\omega}_i^2$ was computed using the formula $\hat{\theta}_i^2$ / (ΣMSF$_i$ + MSE$_b$ + MSE$_w$), where MSE$_b$ is the between-subjects mean square error and MSE$_w$ is the within-subjects mean square error. The $\hat{\omega}^2$ values were used to select significant effects for interpretation. Only effects for which $\hat{\omega}^2$ was larger than .05 were selected. This criterion is arbitrary. However, inspection of effects with $\hat{\omega}^2$ in the range (.01, .05) suggests these effects are small in magnitude.

Type I Error Rates

Completely True Null Hypothesis

Distributions of $\hat{\tau}$ for the 11 tests are summarized in Table 4. The standard error of these estimated Type I error

Table 4

Distributions of Type I Error Rates for the Eleven Tests: Completely True Null Hypothesis and α = .05

Test	Min	10	25	50	75	90	Max
B	.0234	.0250	.0278	.0374	.0412	.0450	.0472
B_{RT}	.0222	.0260	.0292	.0326	.0346	.0368	.0408
B_{DT}	.0330	.0380	.0406	.0450	.0496	.0544	.0656
W	.0388	.0448	.0482	.0544	.0594	.0656	.0744
W_{RT}	.0360	.0418	.0450	.0484	.0528	.0576	.0664
$S(\tilde{\varepsilon})$	.0030	.0066	.0204	.0326	.0422	.0554	.0762
$S(\tilde{\varepsilon})_{RT}$	.0038	.0090	.0254	.0332	.0378	.0490	.0596
$S(\tilde{\varepsilon})_{DT}$	.0034	.0066	.0190	.0252	.0306	.0498	.0668
S(C)	.0188	.0236	.0272	.0354	.0402	.0448	.0488
$S(C)_{RT}$	.0176	.0264	.0310	.0332	.0354	.0384	.0428
$S(C)_{DT}$	.0154	.0190	.0222	.0256	.0270	.0294	.0348

Note. B = Bonferroni; B_{RT} = raw score trimmed Bonferroni; B_{DT} = trimmed difference score Bonferroni; W = Welsch's step-up procedure; W_{RT} = raw score trimmed Welsch step-up procedure; $S(\varepsilon)$ = Shaffer's Modified Sequentially Rejective Bonferroni (MSRB) procedure using an ε-adjusted omnibus test; $S(\varepsilon)_{RT}$ = Shaffer's raw score trimmed MSRB procedure using an ε-adjusted omnibus test; $S(\varepsilon)_{DT}$ = Shaffer's trimmed difference score MSRB using an ε-adjusted omnibus test; S(C) = Shaffer's Modified Sequentially Rejective Bonferroni (MSRB) procedure using a CIGA omnibus test; $S(C)_{RT}$ = Shaffer's raw score trimmed MSRB procedure using a CIGA omnibus test; $S(C)_{DT}$ = Shaffer's trimmed difference score MSRB using a CIGA omnibus test.

rates is $[\alpha(1-\alpha)/5000]^{1/2}$. If the population Type I error rate were .05, the standard error would be .0031, so that the rejection region for an upper-tailed z test of H_0: $\alpha = .05$ is .055 at a .05 significance level. As would be expected from the theoretical developments underlying the B and S(C) tests, using the upper-tailed z test these tests do not result in Type I error rates above .05. The B_{RT}, $S(C)_{RT}$, and $S(C)_{DT}$ tests also share the property and, in fact, appear to be more conservative than their non-trimmed counterparts. As expected from theory and empirical results, the $S(\tilde{\varepsilon})$ test can be liberal in some conditions. However, the test is not too badly liberal and trimming tempers the liberal tendency. The Welsch test appears to have the strongest liberal tendency, presumably because the critical value for the Welsch test has no theoretical support when multisample sphericity is violated. In the trimmed version, W_{RT}, the liberal tendency is reduced. By Bradley's (1978) liberal criterion a test is robust if $.5\alpha \leq \tau \leq 1.5\alpha$, where α is the nominal significance level and τ is the true probability of Type I error rate. None of the tests are liberal using Bradley's criterion.

Selecting among the tests would depend on one's tolerance for tests that can be somewhat liberal in some conditions. To avoid such tests and to avoid low Type I error rates, B appears to be the best choice. If a somewhat liberal test can be tolerated B_{DT}, W or W_{RT} appear to be the best choice.

A 3 (Distribution) x 2 (Sphericity) x 3 (V-C Heteroscedasticity) x 3 (Sample Size Arrangement) x 2 (Total Sample Size) x 11 (Test) ANOVA, with repeated measures on the test factor was used to analyze the Type I error rates ($\hat{\tau}$). Among the 30 effects involving only between-subjects factors, the main effects of D, n_1/n_2, and N, and the n_1/n_2 x $\Sigma_1:\Sigma_2$ interaction were significant. Among the 31 effects involving the T factor and interactions of test with the between-subjects factors, the main effect of T, all 5 two-way interactions, and 9 higher order interactions were significant. Shown in Table 5 are the effects that accounted for more than 1% of the total variance.

Effect of T. Shown in Table 6 are mean values of $\hat{\tau}$ for the eleven tests. The means are similar to the medians in Table 4 and do not alter the interpretations presented in the discussion of Table 4.

The T x n_1/n_2 x $\Sigma_1:\Sigma_2$ and T x n_1/n_2 effects. Means for interpreting the T x n_1/n_2 x $\Sigma_1:\Sigma_2$ interaction are presented in Table 7. Results in the table indicate that when $n_1 = n_2$, heterogeneity of the v-c matrices has little effect on $\hat{\tau}$. Based on Type I error rates, when $n_1 = n_2$, the best procedures appear to be W, W_{RT}, and B_{DT} because they maintain $\hat{\tau}$ nearest to α. When $\Sigma_1:\Sigma_2 = 1:1$, n_1/n_2 appears to have relatively little effect on $\hat{\tau}$. However, when the design is unbalanced, those concerned about liberal tests might avoid W and use either W_{RT} or B_{DT}. When heterogeneous v-c matrices are

Table 5

Percent of Variance for Type I Error Rates for Effects that Accounted for at least 1% of the variance: Completely True Null Hypothesis and $\alpha = .05$

Effect	$\hat{\omega}^2$
T	0.623
T x n_1/n_2	0.178
T x n_1/n_2 x $\Sigma_1:\Sigma_2$	0.099
T x D	0.023
D	0.013
T x ε	0.013

Table 6

Mean Type I Error Rates as a Function of Test: Completely True Null Hypothesis and $\alpha = .05$

Test	mean $\hat{\tau}$
B	0.035
B_{RT}	0.032
B_{DT}	0.046
W	0.055
W_{RT}	0.049
$S(\tilde{\varepsilon})$	0.033
$S(\tilde{\varepsilon})_{RT}$	0.032
$S(\tilde{\varepsilon})_{DT}$	0.027
S(C)	0.035
$S(C)_{RT}$	0.033
$S(C)_{DT}$	0.025

Note. See Table 4 for abbreviations.

Table 7

Mean Type I Error Rates as a Function of V-C Heterogeneity, Test, and Sample Size Arrangement: Completely True Null Hypothesis and $\alpha = .05$

$\Sigma_1:\Sigma_2$	Test	$n_1<n_2$	$n_1=n_2$	$n_1>n_2$
1:1	B	0.037	0.036	0.035
	B_{RT}	0.033	0.031	0.032
	B_{DT}	0.048	0.043	0.046
	W	0.058	0.053	0.054
	W_{RT}	0.051	0.046	0.050
	$S(\tilde{\varepsilon})$	0.032	0.036	0.031
	$S(\tilde{\varepsilon})_{RT}$	0.032	0.034	0.031
	$S(\tilde{\varepsilon})_{DT}$	0.027	0.024	0.026
	S(C)	0.035	0.036	0.034
	$S(C)_{RT}$	0.033	0.034	0.033
	$S(C)_{DT}$	0.026	0.024	0.025

Table 7 -- continued.

$\Sigma_1:\Sigma_2$	Test	$n_1<n_2$	$n_1=n_2$	$n_1>n_2$
1:2	B	0.035	0.035	0.036
	B_{RT}	0.032	0.030	0.033
	B_{DT}	0.044	0.043	0.050
	W	0.054	0.053	0.056
	W_{RT}	0.048	0.046	0.053
	$S(\tilde{\varepsilon})$	0.016	0.036	0.047
	$S(\tilde{\varepsilon})_{RT}$	0.017	0.034	0.045
	$S(\tilde{\varepsilon})_{DT}$	0.014	0.024	0.041
	S(C)	0.036	0.035	0.033
	$S(C)_{RT}$	0.033	0.034	0.032
	$S(C)_{DT}$	0.025	0.024	0.026

Table 7 -- continued.

$\Sigma_1:\Sigma_2$	Test	$n_1<n_2$	$n_1=n_2$	$n_1>n_2$
1:5	B	0.034	0.035	0.037
	B_{RT}	0.032	0.032	0.032
	B_{DT}	0.042	0.045	0.052
	W	0.052	0.053	0.059
	W_{RT}	0.047	0.048	0.054
	$S(\tilde{\varepsilon})$	0.005	0.035	0.058
	$S(\tilde{\varepsilon})_{RT}$	0.007	0.034	0.052
	$S(\tilde{\varepsilon})_{DT}$	0.006	0.025	0.056
	S(C)	0.035	0.034	0.033
	$S(C)_{RT}$	0.034	0.033	0.029
	$S(C)_{DT}$	0.025	0.024	0.026

Note. See Table 4 for abbreviations.

positively paired with unequal group sizes ($\Sigma_1:\Sigma_2 \neq 1:1$, $n_1 < n_2$), Type I error rates are most strongly deflated for $S(\tilde{\varepsilon})$, $S(\tilde{\varepsilon})_{RT}$, and $S(\tilde{\varepsilon})_{DT}$. The Bonferroni and Welsch tests show a minor degree of inflation. The $S(C)$, $S(C)_{RT}$, and $S(C)_{DT}$ tests show no deflation. When heterogeneous v-c matrices are negatively paired with unequal group sizes ($\Sigma_1:\Sigma_2 \neq 1:1$, $n_1 > n_2$), Type I error rates are most strongly inflated for the $S(\tilde{\varepsilon})$, $S(\tilde{\varepsilon})_{RT}$, and $S(\tilde{\varepsilon})_{DT}$ tests. The B, B_{RT}, $S(C)$, $S(C)_{RT}$,$S(C)_{DT}$ exhibit little or no inflation. The remaining tests (B_{DT}, W, and W_{RT}) exhibit moderate inflation. If heterogeneity of v-c matrices is expected, those concerned about liberal tests may want to use B_{DT} or W_{RT} rather than W. To interpret T x n_1/n_2 effect the means in Table 7 would be collapsed over values of $\Sigma_1:\Sigma_2$. This would yield means similar to those for $\Sigma_1:\Sigma_2 = 1:2$. Consequently the T x n_1/n_2 effect does not provide any new information about the effect of T and n_1/n_2 on $\hat{\tau}$.

Partially True Null Hypothesis

Distributions of $\hat{\tau}$ for partially true null hypotheses are summarized in Table 8 for the 11 tests. Because all of the tests are designed to control the familywise Type I error rate for a completely true null hypothesis, they have very low Type I error rates for a partially true null hypothesis. The procedures using trimmed data appear to be more conservative than their non-trimmed counterparts. The W and W_{RT} tests have Type I error rates closest to the nominal significance level.

Table 8

Distributions of Type I Error Rates for the Eleven Tests: Partially True Null Hypothesis and α = .05

Test	Min	10	25	50	75	90	Max
B	.0028	.0052	.0072	.0096	.0150	.0174	.0206
B_{RT}	.0036	.0050	.0060	.0086	.0128	.0148	.0180
B_{DT}	.0052	.0070	.0086	.0116	.0168	.0200	.0252
W	.0082	.0122	.0174	.0234	.0374	.0428	.0488
W_{RT}	.0092	.0122	.0148	.0202	.0328	.0364	.0438
$S(\tilde{\varepsilon})$	.0012	.0046	.0090	.0138	.0218	.0292	.0380
$S(\tilde{\varepsilon})_{RT}$	.0024	.0062	.0090	.0130	.0204	.0258	.0326
$S(\tilde{\varepsilon})_{DT}$	.0028	.0062	.0098	.0142	.0212	.0256	.0386
S(C)	.0022	.0052	.0096	.0140	.0238	.0284	.0340
$S(C)_{RT}$	.0030	.0064	.0088	.0144	.0208	.0236	.0302
$S(C)_{DT}$	.0040	.0068	.0094	.0140	.0214	.0248	.0328

Note. See Table 4 for abbreviations.

A 2 (Configuration of Means) x 3 (Distribution) x 2 (Sphericity) x 3 (V-C Heteroscedasticity) x 3 (Sample Size Arrangement) x 2 (Total Sample Size) x 11 (Test) ANOVA, with repeated measures on the test factor was used to analyze the $\hat{\tau}$s. The two configurations of means for which the null hypothesis is partially true are the minimum range and the maximum range configurations. Among the 62 effects involving only between-subjects factor, all main effects except v-c heteroscedasticity, 4 two-way interactions, and 2 higher order interactions were significant. Among the 63 effects involving the T factor and interactions of test with the between-subjects factors, the main effect of T, all 5 two-way interactions, and 20 higher order interactions were significant. Shown in Table 9 are the effects that accounted for more than 1% of the total variance.

The T, CM, and T x CM effects. Shown in Table 10 are mean values of $\hat{\tau}$ for the 11 tests. The means are similar to the medians in Table 8 and do not alter the interpretations presented in the discussion of Table 8. Means for interpreting the T x CM interaction are presented in Table 11. Results in the table show that Type I error rates are higher for the minimum range configuration than for the maximum range configuration. This result occurs because the minimum range configuration has two true pairwise comparisons and the maximum range configuration has one true pairwise comparison.

Table 9

Percent of Variance for Type I error rates for Effects that Accounted for at least 1% of the Variance: Partially True Null Hypothesis and $\alpha = .05$

Effect	$\hat{\omega}^2$
T	0.618
T x CM	0.088
CM	0.085
T x D	0.061
T x n_1/n_2	0.039
T x n_1/n_2 x $\Sigma_1:\Sigma_2$	0.024
D	0.022
T x ε	0.011

Table 10

Mean Type I error rates as a Function of Test: Partially True Null Hypothesis and $\alpha = .05$

Test	mean $\hat{\tau}$
B	0.011
B_{RT}	0.009
B_{DT}	0.013
W	0.026
W_{RT}	0.024
$S(\tilde{\varepsilon})$	0.016
$S(\tilde{\varepsilon})_{RT}$	0.015
$S(\tilde{\varepsilon})_{DT}$	0.016
S(C)	0.016
$S(C)_{RT}$	0.015
$S(C)_{DT}$	0.015

Note. See Table 4 for abbreviations.

Table 11

Mean Type I error rates as a Function of Test and the Configuration of Means: Partially True Null Hypothesis and α = .05

Test	Minimum range	Maximum range
B	0.014	0.007
B_{RT}	0.013	0.006
B_{DT}	0.017	0.009
W	0.035	0.017
W_{RT}	0.032	0.015
$S(\tilde{\varepsilon})$	0.021	0.010
$S(\tilde{\varepsilon})_{RT}$	0.020	0.009
$S(\tilde{\varepsilon})_{DT}$	0.021	0.010
S(C)	0.022	0.010
$S(C)_{RT}$	0.020	0.009
$S(C)_{DT}$	0.021	0.010

Note. See Table 4 for abbreviations.

For both configurations, Type I error rates are closest to .05 for the W and W_{RT} tests.

The T x D effect. In Table 12, mean Type I error rates are reported for each distribution. All test procedures have lower Type I error rates when the distribution has long tails. As kurtosis increases the Type I error rates for the procedures using non-trimmed data tend to decline more rapidly than do the Type I error rates for procedures using trimmed data. Among the 11 tests, the W and W_{RT} tests tend to have Type I error rates closest to .05. For a normal distribution, W has the Type I error rate that is closest to .05. As kurtosis increases the Type I error rates for W and W_{RT} become more similar.

Power

In this section power results are presented in three subsections: any-pair power, average per-pair power, and all-pairs power. In each subsection, percentiles of the distribution of power are presented for each test. Then the results of a 3 (Configuration of Means) x 3 (Distribution) x 2 (Sphericity) x 3 (V-C Heteroscedasticity) x 3 (Sample Size Arrangement) x 2 (Total Sample Size) x 11 (Test) ANOVA, with repeated measures on the test factor are presented.

Any-Pair Power

Distributions of any-pair power for the 11 tests are summarized in Table 13. The W and W_{RT} tests have higher any-

Table 12

Mean Type I error rates as a Function of Test and Distribution: Partially True Null Hypothesis

Test	h=0	h=.109	h=.35
B	0.013	0.012	0.008
B_{RT}	0.011	0.009	0.008
B_{DT}	0.014	0.013	0.011
W	0.032	0.028	0.019
W_{RT}	0.027	0.024	0.020
$S(\tilde{\varepsilon})$	0.021	0.017	0.009
$S(\tilde{\varepsilon})_{RT}$	0.017	0.015	0.012
$S(\tilde{\varepsilon})_{DT}$	0.018	0.016	0.012
S(C)	0.021	0.018	0.009
$S(C)_{RT}$	0.017	0.015	0.012
$S(C)_{DT}$	0.018	0.016	0.011

Note. See Table 4 for abbreviations.

Table 13

Distributions of Any-Pair Power for the Eleven Tests at $\alpha = .05$

Test	Min	10	25	50	75	90	Max
B	.078	.103	.130	.269	.361	.443	.699
B_{RT}	.100	.146	.178	.222	.282	.341	.541
B_{DT}	.126	.164	.203	.273	.339	.402	.621
W	.123	.147	.175	.327	.425	.505	.777
W_{RT}	.152	.201	.232	.290	.346	.410	.633
$S(\tilde{\varepsilon})$	.032	.088	.110	.260	.337	.432	.486
$S(\tilde{\varepsilon})_{RT}$	.094	.165	.188	.246	.288	.315	.368
$S(\tilde{\varepsilon})_{DT}$	.085	.145	.167	.239	.280	.316	.377
S(C)	.077	.097	.111	.274	.361	.450	.629
$S(C)_{RT}$	.114	.168	.196	.253	.301	.348	.512
$S(C)_{DT}$	.102	.144	.166	.238	.289	.335	.481

Note. See Table 4 for abbreviations.

pair power than the other tests, and thus these results favor the use of the W and W_{RT} tests. However these tests, along with B_{DT} had some tendency to be liberal when the null hypothesis was completely true. Based on the results in Table 13, those who want to avoid a liberal test should select the B or S(C) test. Among the 62 effects involving only between-subjects factors, all main effects except v-c heteroscedasticity, 9 two-way interactions, and 7 higher order interactions were significant. Among the 63 effects involvingthe T factor and interactions of test with the between-subjects factors, the main effect of T, all 6 two-way interactions, and 34 higher order interactions were significant. Shown in Table 14 are the effects that accounted for more than 1% of the total variance.

The T, D, and T x D effects. Means of any-pair power rates for interpreting the T x D interaction are presented in Table 15. Results in the table indicate that all eleven tests have less power when the data are generated from long-tailed distributions. Tests using non-trimmed data have high power when the distribution is normal. The tests using trimmed data have about 10 to 25 percent less power than their non-trimmed counterparts. The W test has the highest any-pair power when the distribution is normal. The W test also has the highest power when the distribution is slightly kurtotic ($h = .109$), but W_{RT}, B, B_{DT}, and S(C) perform well, also. In the very long-tailed non-normal distribution ($h = .35$), W_{RT} and B_{DT} have

Table 14

Percent of Variance of Any-Pair Power for Effects that Accounted for at least 1% of the Variance: $\alpha = .05$

Effect	$\hat{\omega}^2$
T x D	0.234
T	0.200
T x n_1/n_2	0.138
D	0.132
T x n_1/n_2 x $\Sigma_1:\Sigma_2$	0.094
T x CM x ε	0.058
T x CM	0.024
T x ε	0.018
T x $\Sigma_1:\Sigma_2$	0.013
T x n_1/n_2 x D	0.012
CM x ε	0.010

Table 15

Mean of Any-Pair Power as a Function of Test and Distribution: $\alpha = .05$

Test	h=0	h=.109	h=.35
B	0.396	0.289	0.117
B_{RD}	0.288	0.249	0.171
B_{DT}	0.359	0.296	0.189
W	0.468	0.355	0.165
W_{RT}	0.360	0.316	0.226
$S(\tilde{\varepsilon})$	0.381	0.268	0.093
$S(\tilde{\varepsilon})_{RT}$	0.290	0.252	0.175
$S(\tilde{\varepsilon})_{DT}$	0.292	0.243	0.153
S(C)	0.408	0.291	0.105
$S(C)_{RT}$	0.311	0.270	0.188
$S(C)_{DT}$	0.303	0.251	0.156

Note. See Table 4 for abbreviations.

the highest powers. The procedures using non-trimmed data have low any-pair power when h = .35.

The T x n_1/n_2 and T x n_1/n_2 x $\Sigma_1:\Sigma_2$ effects. Means for interpreting the T x n_1/n_2 x $\Sigma_1:\Sigma_2$ interaction are presented in Table 16. Based on any-pair power, when $n_1 = n_2$, the best procedures appear to be W, W_{RT}, and B_{DT}. However, the W test had slightly elevated Type I error rates when $n_1 = n_2$ and the null hypothesis was completely true. When $\Sigma_1:\Sigma_2 = 1:1$, n_1/n_2 appears to have relatively little effect on any-pair power. When heterogeneous v-c matrices are positively paired with unequal group sizes ($\Sigma_1:\Sigma_2 \neq 1:1$, $n_1 < n_2$), any-pair power rates for Shaffer's procedures using the $\tilde{\varepsilon}$-adjusted omnibus F test decline; power rates for the remaining tests increase. For all tests the degree of deflation or inflation increases with increases in the degree of v-c matrices heterogeneity. When heterogeneous v-c matrices are negatively paired with unequal group sizes ($\Sigma_1:\Sigma_2 \neq 1:1$, $n_1 > n_2$), any-pair power rates for all tests except Shaffer's procedures using the $\tilde{\varepsilon}$-adjusted omnibus F test decline, with the degree of deflation increasing with increases in the degree of v-c matrices heterogeneity. Shaffer's procedures following an $\tilde{\varepsilon}$-adjusted omnibus F test exhibit no deflation. The W and W_{RT} tests have the highest power for both positive and negative pairing; Shaffer's procedures following $\tilde{\varepsilon}$-adjusted F test have very low power in positive pairing.

Table 16

Mean of Any-Pair Power as a Function of V-C Heterogeneity, Test, and Sample Size: α = .05

$\Sigma_1:\Sigma_2$	Test	$n_1<n_2$	$n_1=n_2$	$n_1>n_2$
1:1	B	0.260	0.267	0.261
	B_{RT}	0.229	0.234	0.230
	B_{DT}	0.277	0.277	0.276
	W	0.323	0.323	0.325
	W_{RT}	0.296	0.293	0.296
	$S(\tilde{\varepsilon})$	0.254	0.271	0.256
	$S(\tilde{\varepsilon})_{RT}$	0.250	0.261	0.251
	$S(\tilde{\varepsilon})_{DT}$	0.239	0.235	0.239
	S(C)	0.262	0.268	0.263
	$S(C)_{RT}$	0.249	0.259	0.250
	$S(C)_{DT}$	0.233	0.233	0.233

Table 16 -- continued.

$\Sigma_1:\Sigma_2$	Test	$n_1<n_2$	$n_1=n_2$	$n_1>n_2$
1:2	B	0.302	0.265	0.229
	B_{RT}	0.272	0.232	0.196
	B_{DT}	0.314	0.275	0.245
	W	0.363	0.322	0.294
	W_{RT}	0.339	0.292	0.263
	$S(\tilde{\varepsilon})$	0.214	0.269	0.272
	$S(\tilde{\varepsilon})_{RT}$	0.219	0.258	0.254
	$S(\tilde{\varepsilon})_{DT}$	0.203	0.233	0.260
	S(C)	0.303	0.266	0.229
	$S(C)_{RT}$	0.296	0.255	0.211
	$S(C)_{DT}$	0.274	0.230	0.199

Table 16 -- continued.

$\Sigma_1:\Sigma_2$	Test	$n_1<n_2$	$n_1=n_2$	$n_1>n_2$
1:5	B	0.357	0.265	0.203
	B_{RT}	0.334	0.229	0.166
	B_{DT}	0.374	0.278	0.218
	W	0.419	0.326	0.268
	W_{RT}	0.401	0.293	0.232
	$S(\tilde{\varepsilon})$	0.155	0.268	0.270
	$S(\tilde{\varepsilon})_{RT}$	0.166	0.254	0.238
	$S(\tilde{\varepsilon})_{DT}$	0.158	0.232	0.265
	S(C)	0.358	0.264	0.199
	$S(C)_{RT}$	0.362	0.249	0.177
	$S(C)_{DT}$	0.335	0.226	0.170

Note. See Table 4 for abbreviations.

The T x CM x ε effect. Means for interpreting the T x CM x ε interaction are presented in Table 17. For all combinations of configurations and ε, W has the highest any-pair power, followed by W_{RT}. The results in Table 17 indicate that (a) the impact of mean configuration on power is smaller for the Shaffer tests than it is for the Welsch and Bonferroni tests and (b) the impact of mean configuration on power depends on ε. With respect to (b) when $\varepsilon = .96$, power is highest for the maximum range configuration; when $\varepsilon = .75$, power is highest for the minimum range configuration. The variation in the relationship between mean configuration and power as a function of ε is due to the impact of

$$\max_{k,k'} \frac{\overline{\mu}_k^* - \overline{\mu}_{k'}^*}{\sqrt{\sum_j (s^2_{jk} + s^2_{jk'} - 2s_{jkk'})/n_j}} \tag{52}$$

on power. Because of the means and dispersion matrices, used in simulating the data, when $\varepsilon = .96$ the ratio in equation (52) is largest for the maximum range configuration and smallest for the minimum range configuration. When $\varepsilon = .75$, the ratio in (52) is largest for the minimum range configuration and smallest for the maximum range configuration. The smaller impact of mean configuration on power for the Shaffer tests is due to the impact of the omnibus test on the power of these procedures. For the minimum range configuration any-pair power for the Bonferroni and Welsch tests increases as ε changes from .96 to .75. However, any-pair powers for these tests decrease as ε changes

Table 17

Mean of Any-Pair Power as a Function of Sphericity, Test, and the Configuration of Means: $\alpha = .05$

Sphericity	Test	minimum range	maximum range	equally spaced
.96	B	0.251	0.299	0.278
	B_{RT}	0.219	0.261	0.245
	B_{DT}	0.265	0.306	0.289
	W	0.301	0.347	0.329
	W_{RT}	0.274	0.316	0.300
	$S(\tilde{\varepsilon})$	0.260	0.276	0.270
	$S(\tilde{\varepsilon})_{RT}$	0.244	0.260	0.254
	$S(\tilde{\varepsilon})_{DT}$	0.233	0.245	0.241
	S(C)	0.284	0.301	0.295
	$S(C)_{RT}$	0.265	0.282	0.276
	$S(C)_{DT}$	0.243	0.254	0.249

Table 17 -- continued.

Sphericity	Test	minimum range	maximum range	equally spaced
.75	B	0.351	0.201	0.226
	B_{RT}	0.297	0.187	0.206
	B_{DT}	0.364	0.219	0.246
	W	0.433	0.265	0.301
	W_{RT}	0.381	0.252	0.280
	$S(\tilde{\varepsilon})$	0.242	0.217	0.220
	$S(\tilde{\varepsilon})_{RT}$	0.238	0.216	0.221
	$S(\tilde{\varepsilon})_{DT}$	0.236	0.208	0.215
	S(C)	0.255	0.235	0.238
	$S(C)_{RT}$	0.249	0.231	0.235
	$S(C)_{DT}$	0.240	0.215	0.221

Note. See Table 4 for abbreviations.

from .96 to .75 for the maximum range and equally spaced configurations. These results reflect changes in the ratio in equation (52).

Average Per-Pair Power

Distributions of average per-pair power for the 11 tests are summarized in Table 18. The W and W_{RT} tests have the highest average per-pair power. The Bonferroni procedures have low average per-pair power.

Among the 62 effects involving only between-subjects factors, all 6 main effects, 10 two-way interactions, and 6 higher order interactions were significant at $\alpha = .05$. Among the 63 effects involving the T factor and interactions of test with the between-subjects factors, the main effect of T, all 6 two-way interactions with test, and 32 higher order interactions were significant. The effects that were significant were almost same as the significant effects for any-pair power. Shown in Table 19 are the effects that accounted for more than 1% of the total variance.

The T, T x CM, and T x CM x ε effects. Means for interpreting the T x CM x ε interaction are presented in Table 20. For all combinations of configurations and ε, W has the highest average per-pair power, followed by W_{RT}. As did the results for any-pair power, the results in table indicate that the impact of mean configuration on power is smaller for the Shaffer tests than it is for the Welsch and Bonferroni tests. For both $\varepsilon = .96$ and .75 power is largest for the minimum

Table 18

Distributions of Average Per-Pair Power for the Eleven Tests at $\alpha = .05$

Test	Min	10	25	50	75	90	Max
B	.020	.025	.036	.074	.111	.150	.308
B_{RT}	.024	.038	.048	.064	.085	.111	.218
B_{DT}	.029	.040	.055	.076	.102	.133	.253
W	.047	.064	.098	.170	.249	.396	.723
W_{RT}	.056	.085	.112	.151	.208	.316	.566
$S(\tilde{\varepsilon})$	.014	.029	.036	.098	.134	.177	.220
$S(\tilde{\varepsilon})_{RT}$	.037	.053	.066	.083	.105	.126	.154
$S(\tilde{\varepsilon})_{DT}$	.028	.045	.058	.083	.106	.130	.156
S(C)	.023	.031	.040	.099	.139	.198	.317
$S(C)_{RT}$	.030	.051	.066	.086	.114	.138	.235
$S(C)_{DT}$	.026	.044	.059	.082	.110	.139	.242

Note. See Table 4 for abbreviations.

Table 19

Percent of Variance for Average Per-Pair Power for Effects that Accounted for at least 1% of the Variance: $\alpha = .05$

Effect	$\hat{\omega}^2$
T	0.509
T x CM	0.134
T x D	0.113
T x CM x ε	0.050
D	0.040
T x ε	0.032
CM	0.021
T x n_1/n_2	0.021
T x CM x D	0.020
T x n_1/n_2 x $\Sigma_1:\Sigma_2$	0.015

Table 20

Mean of Average Per-Pair Power as a Function of Sphericity, Test, and the Configuration of Means: $\alpha = .05$

Sphericity	Test	minimum range	maximum range	equally spaced
.96	B	0.097	0.085	0.068
	B_{RT}	0.079	0.071	0.057
	B_{DT}	0.094	0.083	0.067
	W	0.214	0.189	0.111
	W_{RT}	0.189	0.166	0.098
	$S(\tilde{\varepsilon})$	0.120	0.096	0.081
	$S(\tilde{\varepsilon})_{RT}$	0.105	0.085	0.072
	$S(\tilde{\varepsilon})_{DT}$	0.102	0.081	0.069
	S(C)	0.129	0.104	0.088
	$S(C)_{RT}$	0.112	0.092	0.077
	$S(C)_{DT}$	0.107	0.085	0.073

Table 20 -- continued.

Sphericity	Test	minimum range	maximum range	equally spaced
.75	B	0.130	0.067	0.061
	B_{RT}	0.105	0.058	0.052
	B_{DT}	0.127	0.066	0.061
	W	0.377	0.176	0.121
	W_{RT}	0.319	0.161	0.107
	$S(\tilde{\varepsilon})$	0.121	0.087	0.075
	$S(\tilde{\varepsilon})_{RT}$	0.108	0.080	0.069
	$S(\tilde{\varepsilon})_{DT}$	0.108	0.078	0.068
	S(C)	0.128	0.093	0.080
	$S(C)_{RT}$	0.114	0.084	0.073
	$S(C)_{DT}$	0.113	0.081	0.071

Note. See Table 4 for abbreviation.

range configuration and smallest for the equally spaced configuration. The effect of mean configuration is due to the magnitude of the differences between nonnull means in the three types of configurations. For the minimum range configuration, the four differences between nonnull means have same size. For the maximum range configuration and the equally spaced configuration, the difference between the largest and smallest means is large, but the other four differences are moderate in comparison to the difference for the minimum range configuration. For the minimum range configuration, average per-pair power tends to increase for all tests as ε changes from .96 to .75. The largest increases are for the W and W_{RT} tests. For all tests the effect of ε on power is small for the maximum range and equally spaced mean configurations. This pattern of results is predictable from the average value of

$$\frac{\bar{\mu}_{k}^{*} - \bar{\mu}_{k'}^{*}}{\sqrt{\sum_{j} (s^{*}_{jk} + s^{*}_{jk'} - 2s_{jkk'})/n_{j}}} \tag{53}$$

for the non-null contrasts for each combination of ε and mean configuration. As ε changes from .96 to .75, the change in this average is larger for the minimum range configuration, than it is for the other two configurations.

The T x D effect. Means for interpreting the T x D interaction are presented in Table 21. The T x D interaction for average per-pair power is similar to the T x D interaction for any-pair power. The results in the Table 21 indicate that

Table 21

Mean of Average Per-Pair Power as a Function of Test and Distribution: α = .05

Test	h=0	h=.109	h=.35
B	0.133	0.090	0.032
B_{RT}	0.089	0.074	0.048
B_{DT}	0.110	0.087	0.052
W	0.289	0.213	0.093
W_{RT}	0.212	0.182	0.126
$S(\tilde{\varepsilon})$	0.156	0.103	0.032
$S(\tilde{\varepsilon})_{RT}$	0.109	0.091	0.060
$S(\tilde{\varepsilon})_{DT}$	0.113	0.089	0.051
S(C)	0.165	0.111	0.036
$S(C)_{RT}$	0.115	0.097	0.064
$S(C)_{DT}$	0.118	0.093	0.053

Note. See Table 4 for abbreviations.

all eleven tests have less power when the data are generated from long-tailed distributions. The W test has the highest average per-pair power in the normal distribution. The tests using non-trimmed data have high power when the distribution is normal. The tests using trimmed data have about 25 to 40% less power than their counterparts. The W test has the highest power when h = .109, and the trimmed version, W_{RT}, has the next highest power. In the very long-tailed distribution (h = .35), W_{RT} has the highest average per-pair power and W has the next highest power.

All-Pairs Power

Distribution of all-pairs power for the 11 tests are summarized in Table 22. All eleven tests have low all-pairs power. However, power is highest for the W and W_{RT} tests. Because power is zero for so many cells, the results of the ANOVA for the full set of the data will not be reported. The $\hat{\omega}^2$ values were calculated and suggest a strong T x CM interaction. Means for interpreting the interaction are presented in Table 23. The results indicate power is very low except for the W and W_{RT} tests under the minimum range configuration.

For the W and W_{RT} tests under the minimum range configuration, a 3 (Distribution) x 2 (Sphericity) x 3 (V-C Heteroscedasticity) x 3 (Sample Size Arrangement) x 2 (Total Sample Size) x 2 (Test) ANOVA, with repeated measures on the test factor was conducted. Among the 30 effects involving

Table 22

Distributions of All-Pairs Power for the Eleven Tests at $\alpha = .05$

Test	Min	10	25	50	75	90	Max
B	0	0	0	0	.0018	.0122	.0476
B_{RT}	0	0	0	0	.0020	.0060	.0228
B_{DT}	0	0	0	0	.0018	.0062	.0250
W	0	0	0	.0096	.1216	.3218	.6622
W_{RT}	0	0	0	.0064	.1222	.2384	.4940
$S(\tilde{\varepsilon})$	0	0	0	.0004	.0068	.0272	.0844
$S(\tilde{\varepsilon})_{RT}$	0	0	0	.0004	.0060	.0154	.0496
$S(\tilde{\varepsilon})_{DT}$	0	0	0	.0006	.0060	.0160	.0472
$S(C)$	0	0	0	.0004	.0068	.0272	.0864
$S(C)_{RT}$	0	0	0	.0004	.0060	.0154	.0498
$S(C)_{DT}$	0	0	0	.0006	.0064	.0160	.0528

Note. See Table 4 for abbreviations.

Table 23

Mean of All-Pairs Power as a Function of Test and the Configuration of Means: $\alpha = .05$

Test	minimum range	maximum range	equally spaced
B	.00941	.00036	.00000
B_{RT}	.00503	.00015	.00000
B_{DT}	.00521	.00021	.00000
W	.23246	.01151	.00001
W_{RT}	.18933	.00774	.00000
$S(\tilde{\varepsilon})$	.02049	.00200	.00005
$S(\tilde{\varepsilon})_{RT}$	.01291	.00102	.00002
$S(\tilde{\varepsilon})_{DT}$	.01256	.00135	.00009
S(C)	.02059	.00200	.00005
$S(C)_{RT}$	.01292	.00102	.00002
$S(C)_{DT}$	.01293	.00136	.00009

Note. See Table 4 for abbreviations.

only between-subjects factors, all 5 main effects, 5 two-way interactions, and 3 higher order interactions were significant. Among the 31 effects involving the T factor and interactions of test with the between-subjects factors, the main effect of T, 4 two-way interactions, and 6 higher order interactions were significant. Shown in Table 24 are the effects that accounted for more than 1% of the total variance.

Effect of ε. Means for interpreting the effect of ε are presented in Table 25. The all-pairs power for the W and W_{RT} tests are higher when $\varepsilon = .75$ than when $\varepsilon = .96$ for the minimum range configuration. The effect of ε is due to changes in the average value of equation (53) when ε changes from .96 to .75. Because of means and v-c matrices selected for this study, the average value of equation (53) increases when ε changes from .96 to .75.

The T, D, and T x D effects. Means for interpreting the T x D interaction are presented in Table 26. Results in the table indicate that both of the tests have smaller power when the data are generated from long-tailed distributions. As kurtosis increases the power rates for the W decline more rapidly than do the power rates for the W_{RT}. The W test has better power when the distribution is normal ($h = 0$) and slightly kurtotic ($h = .109$). The W_{RT} test has better power when $h = .35$.

Table 24

Percent of Variance for All-Pairs Power for Effects that Accounted for at least 1% of the Variance: $\alpha = .05$

Effect	$\hat{\omega}^2$
ε	0.338
D	0.322
T x D	0.097
T	0.053
n_1/n_2	0.049
D x ε	0.045
n_1/n_2 x $\Sigma_1:\Sigma_2$	0.036
T x ε	0.012
T x D x ε	0.012

Note. Only W and W_{RT} tests are included in ANOVA.

Table 25

Mean All-Pairs Power for Sphericity: $\alpha = .05$

$\varepsilon=.96$	$\varepsilon=.75$
.134	.288

Table 26

Mean of All-Pairs Power as a Function of Test and Distribution: $\alpha = .05$

Test	h=0	h=.109	h=.35
W	.351	.250	.097
W_{RT}	.240	.199	.129

Note. See Table 4 for abbreviations.

CHAPTER 5
DISCUSSION AND CONCLUSIONS

This chapter is divided into three sections: (a) findings of related research, (b) conclusions, and (c) suggestions for future research.

Findings of Related Research

Effect of Non-sphericity

In an investigation of multiple comparison procedures for a design with a single within-subjects factor, Maxwell (1980) found that the Bonferroni procedure became conservative when sphericity assumption was violated. Keselman, Keselman, and Shaffer (1991) studied the split plot design with one between-subjects factor and one within-subjects factor. They used test statistic in equation (3), a test statistic that does not require the assumption of equal v-c matrices. The critical values studied were the Bonferroni, Studentized range, Studentized maximum modulus, and Cochran critical values. Keselman, Keselman, and Shaffer found that the Type I error rates decreased as ε decreased for both normal and nonnormal data regardless of the degree of heteroscedasticity. Keselman (1994) investigated 28 pairwise multiple comparison procedures for comparing repeated measures means in the split plot design with one between-subjects factor and one within-subjects factor. Keselman found that some procedures including Shaffer's modified sequentially rejective Bonferroni procedures using the $\bar{\varepsilon}$-adjusted F or Hotelling's T^2 as omnibus tests became conservative when $\varepsilon = .40$. Welsch's step-up procedure did not become conservative as non-sphericity increased.

Table 5 shows that when the omnibus null hypothesis is completely true, the T x ε interaction accounted for 1.3% of the total variance. Shown in Table 27 are mean values of $\hat{\tau}$ for interpreting the T x ε interaction. These results indicate that when non-sphericity increases, $\hat{\tau}$ decreases slightly for Bonferroni procedures and Shaffer's procedures. Type I error rates for W and W_{RT} increase slightly when non-sphericity increases. The relatively small effect of violating the sphericity assumption might be due to the levels of the non-sphericity included in this study ($\varepsilon = .75$ and $.96$). Other studies have included smaller values of ε.

Effect of Nonnormality

Keselman, Keselman, and Shaffer (1991) studied four simultaneous multiple comparison procedures in the split plot design with one between-subjects factor and one within-subjects factor. They found that the Type I error rates for the Bonferroni procedure applied to data sampled from a chi-square distribution with three degrees of freedom are not consistently larger or smaller than when the Bonferroni

Table 27

Mean Type I Error Rates as a Function of Test and Sphericity: Completely True Null Hypothesis and $\alpha = .05$

Test	ε=.96	ε=.75
B	0.036	0.035
B_{RT}	0.032	0.031
B_{DT}	0.047	0.045
W	0.053	0.057
W_{RT}	0.047	0.051
$S(\tilde{\varepsilon})$	0.035	0.031
$S(\tilde{\varepsilon})_{RT}$	0.033	0.031
$S(\tilde{\varepsilon})_{DT}$	0.027	0.027
S(C)	0.036	0.033
$S(C)_{RT}$	0.034	0.032
$S(C)_{DT}$	0.024	0.025

Note. See Table 4 for abbreviations.

procedure is applied to normal data. Keselman (1994) investigated 28 stepwise and simultaneous multiple comparison procedures in the split plot design with one between-subjects factor and one within-subjects factor. Keselman reported that the Type I error rates of the data sampled from chi-square distribution with three degrees of freedom are quite similar to those of the data sampled from normal distribution for Welsch. The conservative nature of Shaffer's modified sequentially rejective Bonferroni procedures was more evident when the data were nonnormal.

Table 5 shows that when the omnibus null hypothesis is completely true the T x D interaction accounted for 2.3% of the total variance. Shown in Table 28 are mean values of $\hat{\tau}$ for interpreting T x D interaction. Results in the table indicate that Type I error rates are smaller when the data are long-tailed. The explanation for the disagreement between the results in this study and the results in previous research is most likely that this study included more severely long-tailed distribution than did the previous studies. The chi-square distribution with three degrees of freedom has larger kurtosis than the distribution with $h = .109$, but it has much smaller kurtosis than the distribution with $h = .35$. In addition the distributions included in this study were symmetric; the nonnormal distributions in the other studies were asymmetric.

Table 28

Mean Type I Error Rates as a Function of Test and Distribution: Completely True Null Hypothesis and α = .05

Test	h=0	h=.109	h=.35
B	0.042	0.038	0.027
B_{RT}	0.035	0.033	0.028
B_{DT}	0.050	0.047	0.040
W	0.060	0.057	0.047
W_{RT}	0.053	0.050	0.044
$S(\tilde{\varepsilon})$	0.039	0.036	0.023
$S(\tilde{\varepsilon})_{RT}$	0.034	0.033	0.029
$S(\tilde{\varepsilon})_{DT}$	0.030	0.028	0.023
S(C)	0.041	0.038	0.025
$S(C)_{RT}$	0.035	0.034	0.029
$S(C)_{DT}$	0.028	0.026	0.021

Note. See Table 4 for abbreviations.

Effect of Trimmed Means

Yuen (1974) reported that in the case of unequal sample sizes and variances the power of the trimmed Welch test is superior to the power of Welch's test for long-tailed distributions. Wilcox (1992) investigated the effect of nonnormality on Type I error rate and power when comparing two independent groups. Wilcox showed that the power of Welch's procedure drops dramatically as the amount of contamination gets large and Yuen's trimmed Welch test has more power than Welch's test. The results of this study are consistent with the findings of the previous research. Table 15 contains means of any-pair power as a function of test and distribution. Results in the table indicate that the procedures using trimmed data have more any-pair power than their counterparts as the distribution has long tails.

Power Comparisons

Keselman (1994) investigated 28 pairwise multiple comparison procedures in the split plot design with one between-subjects factor and one within-subjects factor. Keselman concluded that the Welsch step-up test tends to have the best any-pair and average per-pair power while the Bonferroni procedure has the least per-pair power. The conclusions of this study is consistent with the findings of the previous research. The results indicate that W and W_{RT} have the highest any-pair and average per-pair power and B, B_{RT}, and B_{DT} have the lowest per-pair power.

Effect of ε on Type I Error Rates for Shaffer's Test

Algina and Oshima (in press) investigated the effect of heterogeneous v-c matrices on the Type I error rates of adjusted F tests using unweighted means in a split plot design. When the sample sizes are equal, the $\tilde{\varepsilon}$-adjusted test provides good control of Type I error rate even when the v-c matrices are heterogeneous. However, when the sample sizes are unequal and v-c matrices are heterogeneous, the $\tilde{\varepsilon}$-adjusted test provides poor control of the Type I error rate. Algina and Oshima concluded that when the sample sizes are not equal, the CIGA test control the Type I error rate much better than does the $\tilde{\varepsilon}$-adjusted test.

The results of this study are consistent with the previous research. Results in Table 7 indicate that when the sample sizes are equal, Shaffer's procedures following $\tilde{\varepsilon}$-adjusted omnibus F tests have good control of Type I error rates even when the v-c matrices are heterogeneous. However, when the sample sizes are unequal and v-c matrices are heterogeneous, Shaffer's procedures following $\tilde{\varepsilon}$-adjusted F test provide poor control of Type I error rates. The poor performance of Shaffer's tests is due to the impact of ε on the Type I error rate of $\tilde{\varepsilon}$-adjusted F test. Shaffer's procedures following CIGA test have good control of Type I error rates under heteroscedasticity when the sample sizes are equal or unequal.

Conclusions

Four hundred and thirty-two simulated conditions were investigated in a complete factorial experiment. The generalization of the results is limited by the range of values for the distribution of the generated data, sphericity of the common v-c matrix, degree of the heteroscedasticity of the v-c matrices, total sample size, sample size arrangement, the configuration of the nonnull means. With these limitations the following conclusions may be set forth:

Conclusion 1. The W and W_{RT} tests have Type I error rates closest to the nominal significance level. However, the W test appears to have the strongest liberal tendency. Especially under conditions where the data are normal and ε is small, those concerned about liberal tests might avoid W and use either W_{RT} or B_{DT}.

Conclusion 2. The W test has the highest power when the distribution is normal or slightly kurtotic. However, W has a liberal tendency in normal distribution as mentioned in Conclusion 1. The W_{RT}, B, B_{DT}, and S(C) tests also perform well in terms of power when the distribution is slightly kurtotic. In the very long-tailed distribution, W_{RT} has the highest powers.

Conclusion 3. The tests using trimmed data have about 10 to 25% less any-pair power and about 25 to 40% less average per-pair power than their non-trimmed counterparts when the distribution is normal.

Conclusion 4. The W and W_{RT} tests have high power under negative pairing of v-c matrices and sample sizes; Shaffer's procedures following $\tilde{\varepsilon}$-adjusted F test have low power under positive pairing. Under negative pairing and large v-c heteroscedasticity the power of the Shaffer tests rival that of W. However, under the cited condition the Shaffer tests tend to be liberal.

Conclusion 5. The W and W_{RT} tests have the highest average per-pair power. The Bonferroni procedures have low average per-pair power.

Conclusion 6. The impact of mean configuration on any-pair power and average per-pair power is smaller for the Shaffer tests than it is for the Welsch and Bonferroni tests.

Suggestions for Future Research

The generalizability of the study is limited by (a) the number of distributions considered, (b) the forms of dispersion matrices considered, (c) the variation in sample size arrangement, and (d) the variation in the power of the omnibus tests. This limitations suggest several lines for additional research.

First, the inclusion of the normal distribution ($h = 0$) and the distributions with $h = .109$ and $h = .35$ gave representation to different levels of kurtosis. Further research needs to be conducted to see how the tests perform and how factors included in the study are affected when (a) data are generated from skewed distributions and (b) data are generated from skewed and long-tailed distributions. In addition short-tailed distributions should be investigated since the Type I error rate for the Welsch test tended to increase as kurtosis declined and to be above α when the data were normal.

Second, additional forms of dispersion matrices should be considered. For example, the autocorrelation (AR) models might be considered. First- and second-order AR models are attractive for the split-plot or repeated measurements designs.

Third, additional permutations of the v-c matrices included in the study should be investigated, since the particular permutations included in this study affected power through the ratio in equation (52).

Fourth, only cases of complete data were considered. Since incomplete data (i.e., data for some conditions are missing) are common in the application of the split-plot or repeated measurements designs, it would be useful to examine the tests when data are incomplete. Also, approaches that does not rely on complete data may be considered. When some subjects have incomplete data, functions can be defined to take the missing data into account (e.g., restriction to those subjects with responses to both conditions for a pairwise difference) or functions that have meaningful definitions for all subjects (e.g., the median) can be considered.

Fifth, in all conditions the power for the omnibus test was approximately .50. However, since all-pairs power was very low for all tests, it would be useful to examine the tests using the noncentrality parameter to produce more power (e.g., omnibus power of .90).

Finally, since nonnormality is common, investigating approaches that do not rely on normality would be useful. Nonparametric methods can be used when the normal distribution assumption is not realistic. Symmetry and independence of the distribution are sufficient to support the sign test for pairwise comparisons, and if the pairwise comparison can also be ranked, the Wilcoxon signed rank test is applicable. The bootstrap provides an alternative approach to assessing the distributional properties of estimators without relying on normality assumption. The bootstrap provides an estimate of the sampling distribution of an estimator based only on the estimation procedure and the sample data in hand without appealing to distributional assumption. The major limitation of the bootstrap in this application may be its heavy use of computer time.

APPENDIX
ESTIMATED TYPE I ERROR RATES AND POWER

Table 29 includes estimated Type I error rates for the trimmed difference score Bonferroni (B_{DT}), Welsch (W), raw score trimmed Welsch (W_{RT}), and Shaffer' procedure following the CIGA test at nominal Type I error rate .05 for differing levels of total sample size, sample size arrangement, distribution, sphericity, and heteroscedasticity of v-c matrices.

Table 30 and Table 31 include estimated any-pair and average per-pair power for the trimmed difference score Bonferroni (B_{DT}), Welsch (W), raw score trimmed Welsch (W_{RT}), and Shaffer' procedure following the CIGA test at nominal Type I error rate .05 for differing levels of configuration of means, total sample size, sample size arrangement, distribution, sphericity, and heteroscedasticity of v-c matrices, respectively. Only results for four most effective procedures are reported. Only any-pair power and average per-pair power are reported since all-pairs power are very low for all procedures.

The levels of the factors are the following: (a) for n_1/n_2, "A" denotes positive pairing of v-c matrices and sample sizes, "B" denotes equal sample sizes, and "C" denotes negative pairing of v-c matrices and sample sizes; (b) for D, "1" denotes the sampled distribution was normal, "2" denotes the sampled distribution was $h = .109$, and "3" denotes the sampled distribution was $h = .35$; (c) for ε, "1" denotes $\varepsilon = .96$ whereas "2" denotes $\varepsilon = .75$; and (d) for $\Sigma_1:\Sigma_2$, "1" denotes $\Sigma_1:\Sigma_2 = 1:1$, "2" denotes $\Sigma_1:\Sigma_2 = 1:2$, and "3" denotes $\Sigma_1:\Sigma_2 = 1:5$. Reported Type I error rates, any-pair power, and average per-pair power are of the form 10000 times estimated rates.

Table 29

Estimated Type I Error Rates: Completely True Null Hypothesis and α = .05

N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
40	A	1	1	1	0612	0664	0552	0442
40	A	1	1	2	0456	0560	0464	0452
40	A	1	1	3	0478	0542	0428	0438
40	A	1	2	1	0580	0692	0596	0376
40	A	1	2	2	0478	0656	0552	0384
40	A	1	2	3	0480	0600	0528	0376
40	A	2	1	1	0544	0646	0502	0420
40	A	2	1	2	0462	0496	0440	0374
40	A	2	1	3	0394	0488	0490	0394
40	A	2	2	1	0476	0684	0540	0358
40	A	2	2	2	0400	0544	0444	0330
40	A	2	2	3	0390	0518	0472	0302
40	A	3	1	1	0406	0440	0436	0250
40	A	3	1	2	0390	0388	0372	0260
40	A	3	1	3	0408	0450	0392	0268
40	A	3	2	1	0360	0446	0450	0226
40	A	3	2	2	0384	0544	0472	0284
40	A	3	2	3	0330	0456	0370	0220
40	B	1	1	1	0430	0528	0418	0448
40	B	1	1	2	0462	0524	0452	0402

Table 29--continued.

N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
40	B	1	1	3	0528	0600	0528	0442
40	B	1	2	1	0476	0652	0562	0416
40	B	1	2	2	0420	0576	0472	0356
40	B	1	2	3	0496	0582	0542	0346
40	B	2	1	1	0488	0520	0458	0418
40	B	2	1	2	0496	0544	0430	0390
40	B	2	1	3	0504	0580	0496	0412
40	B	2	2	1	0388	0574	0456	0314
40	B	2	2	2	0426	0542	0452	0354
40	B	2	2	3	0490	0678	0496	0396
40	B	3	1	1	0380	0442	0360	0272
40	B	3	1	2	0356	0430	0396	0234
40	B	3	1	3	0394	0460	0414	0246
40	B	3	2	1	0408	0526	0406	0252
40	B	3	2	2	0372	0516	0406	0290
40	B	3	2	3	0374	0422	0436	0244
40	C	1	1	1	0506	0580	0490	0392
40	C	1	1	2	0616	0682	0624	0410
40	C	1	1	3	0624	0656	0630	0378
40	C	1	2	1	0506	0636	0546	0404
40	C	1	2	2	0600	0680	0576	0354

Table 29--continued.

N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
40	C	1	2	3	0656	0730	0664	0428
40	C	2	1	1	0476	0512	0524	0352
40	C	2	1	2	0560	0586	0542	0354
40	C	2	1	3	0574	0636	0576	0334
40	C	2	2	1	0490	0638	0530	0384
40	C	2	2	2	0554	0638	0590	0382
40	C	2	2	3	0634	0744	0594	0400
40	C	3	1	1	0440	0456	0460	0230
40	C	3	1	2	0420	0478	0432	0216
40	C	3	1	3	0454	0466	0420	0206
40	C	3	2	1	0428	0496	0478	0236
40	C	3	2	2	0442	0470	0452	0214
40	C	3	2	3	0502	0528	0558	0268
60	A	1	1	1	0528	0548	0502	0404
60	A	1	1	2	0478	0560	0482	0466
60	A	1	1	3	0500	0542	0484	0464
60	A	1	2	1	0492	0700	0576	0418
60	A	1	2	2	0496	0622	0554	0384
60	A	1	2	3	0446	0630	0564	0384
60	A	2	1	1	0452	0588	0474	0434
60	A	2	1	2	0488	0550	0508	0414

Table 29--continued.

N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
60	A	2	1	3	0446	0542	0470	0448
60	A	2	2	1	0430	0528	0502	0312
60	A	2	2	2	0448	0594	0526	0396
60	A	2	2	3	0418	0534	0506	0348
60	A	3	1	1	0422	0456	0494	0236
60	A	3	1	2	0388	0448	0494	0294
60	A	3	1	3	0366	0458	0422	0288
60	A	3	2	1	0402	0546	0476	0278
60	A	3	2	2	0400	0484	0464	0266
60	A	3	2	3	0376	0498	0470	0282
60	B	1	1	1	0466	0538	0478	0488
60	B	1	1	2	0440	0546	0484	0472
60	B	1	1	3	0494	0502	0494	0404
60	B	1	2	1	0460	0578	0542	0324
60	B	1	2	2	0428	0606	0508	0384
60	B	1	2	3	0464	0594	0572	0348
60	B	2	1	1	0462	0546	0462	0458
60	B	2	1	2	0498	0578	0486	0462
60	B	2	1	3	0460	0482	0422	0372
60	B	2	2	1	0412	0562	0484	0322
60	B	2	2	2	0422	0582	0518	0346

Table 29--continued.

N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
60	B	2	2	3	0414	0530	0520	0304
60	B	3	1	1	0390	0430	0416	0266
60	B	3	1	2	0428	0468	0416	0308
60	B	3	1	3	0402	0472	0418	0314
60	B	3	2	1	0388	0506	0472	0290
60	B	3	2	2	0428	0468	0450	0258
60	B	3	2	3	0392	0448	0472	0242
60	C	1	1	1	0534	0586	0492	0458
60	C	1	1	2	0540	0616	0554	0458
60	C	1	1	3	0498	0594	0514	0432
60	C	1	2	1	0448	0608	0526	0370
60	C	1	2	2	0520	0592	0584	0368
60	C	1	2	3	0508	0614	0582	0348
60	C	2	1	1	0450	0536	0480	0398
60	C	2	1	2	0514	0554	0518	0396
60	C	2	1	3	0530	0596	0526	0392
60	C	2	2	1	0438	0560	0536	0364
60	C	2	2	2	0478	0622	0590	0372
60	C	2	2	3	0514	0580	0510	0326
60	C	3	1	1	0372	0422	0444	0246
60	C	3	1	2	0416	0434	0424	0232

Table 29--continued.

N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
60	C	3	1	3	0420	0508	0438	0252
60	C	3	2	1	0388	0472	0454	0244
60	C	3	2	2	0372	0424	0482	0188
60	C	3	2	3	0354	0452	0472	0248

Table 30

Estimated Any-Pair Power: $\alpha = .05$

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
1	40	A	1	1	1	3232	4124	3186	4032
1	40	A	1	1	2	3794	4828	3844	4924
1	40	A	1	1	3	4430	5534	4440	5672
1	40	A	1	2	1	4314	5958	4574	3850
1	40	A	1	2	2	5072	6600	5164	4502
1	40	A	1	2	3	5938	7490	6024	5230
1	40	A	2	1	1	2612	3020	2756	2878
1	40	A	2	1	2	3024	3540	3144	3456
1	40	A	2	1	3	3590	4192	3848	4150
1	40	A	2	2	1	3728	4504	3876	2674
1	40	A	2	2	2	4276	5270	4444	3162
1	40	A	2	2	3	5164	6052	5344	3744
1	40	A	3	1	1	1758	1520	1958	1114
1	40	A	3	1	2	1894	1588	2254	1144
1	40	A	3	1	3	2282	1922	2670	1446
1	40	A	3	2	1	2368	2216	2760	1034
1	40	A	3	2	2	2754	2458	3106	1108
1	40	A	3	2	3	3238	2790	3756	1220
1	40	B	1	1	1	3272	4282	3166	4322
1	40	B	1	1	2	3216	4272	3140	4340

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
1	40	B	1	1	3	3260	4228	3154	4238
1	40	B	1	2	1	4384	6136	4344	4004
1	40	B	1	2	2	4402	6030	4380	3802
1	40	B	1	2	3	4428	6060	4358	3884
1	40	B	2	1	1	2750	3178	2688	3070
1	40	B	2	1	2	2734	3202	2736	3146
1	40	B	2	1	3	2718	3272	2848	3040
1	40	B	2	2	1	3682	4608	3782	2646
1	40	B	2	2	2	3694	4724	3856	2668
1	40	B	2	2	3	3732	4754	3808	2678
1	40	B	3	1	1	1634	1410	1898	1024
1	40	B	3	1	2	1652	1486	1870	1146
1	40	B	3	1	3	1702	1424	1910	1030
1	40	B	3	2	1	2430	2054	2668	936
1	40	B	3	2	2	2386	2126	2754	926
1	40	B	3	2	3	2438	2204	2736	946
1	40	C	1	1	1	3286	4262	3316	4142
1	40	C	1	1	2	2878	3686	2840	3608
1	40	C	1	1	3	2492	3390	2524	3028
1	40	C	1	2	1	4332	5810	4294	3776
1	40	C	1	2	2	3920	5432	3890	3330

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
1	40	C	1	2	3	3374	4782	3386	2822
1	40	C	2	1	1	2688	3194	2744	2916
1	40	C	2	1	2	2400	2850	2424	2472
1	40	C	2	1	3	2080	2562	2128	2176
1	40	C	2	2	1	3606	4686	3932	2762
1	40	C	2	2	2	3108	4088	3350	2294
1	40	C	2	2	3	2906	3754	2846	2026
1	40	C	3	1	1	1652	1566	1892	1098
1	40	C	3	1	2	1590	1484	1778	978
1	40	C	3	1	3	1310	1366	1518	870
1	40	C	3	2	1	2376	2188	2648	940
1	40	C	3	2	2	2060	1984	2320	884
1	40	C	3	2	3	1772	1952	2060	862
1	60	A	1	1	1	3390	4162	3192	4192
1	60	A	1	1	2	3960	4912	3902	5078
1	60	A	1	1	3	4692	5Mw6	4588	5870
1	60	A	1	2	1	4526	5968	4580	3712
1	60	A	1	2	2	5340	6796	5382	4456
1	60	A	1	2	3	6212	7774	6332	5430
1	60	A	2	1	1	2740	3146	2838	3072
1	60	A	2	1	2	3066	3518	3272	3508

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
1	60	A	2	1	3	3866	4170	3966	4284
1	60	A	2	2	1	3888	4688	4002	2754
1	60	A	2	2	2	4504	5248	4624	3100
1	60	A	2	2	3	5206	5968	5436	3754
1	60	A	3	1	1	1884	1452	2070	1126
1	60	A	3	1	2	2042	1586	2304	1264
1	60	A	3	1	3	2402	1696	2720	1366
1	60	A	3	2	1	2418	1982	2754	932
1	60	A	3	2	2	2834	2280	3232	1070
1	60	A	3	2	3	3434	2526	3890	1212
1	60	B	1	1	1	3372	4364	3284	4530
1	60	B	1	1	2	3242	4088	3172	4272
1	60	B	1	1	3	3312	4212	3282	4298
1	60	B	1	2	1	4756	6268	4738	4064
1	60	B	1	2	2	4740	6286	4740	3980
1	60	B	1	2	3	4692	6128	4604	3928
1	60	B	2	1	1	2878	3154	2962	3266
1	60	B	2	1	2	2780	3120	2828	3110
1	60	B	2	1	3	2774	3222	2814	3116
1	60	B	2	2	1	3946	4722	4082	2660
1	60	B	2	2	2	3780	4604	3968	2630

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
1	60	B	2	2	3	3870	4654	4014	2634
1	60	B	3	1	1	1824	1332	2072	1040
1	60	B	3	1	2	1798	1366	1978	1054
1	60	B	3	1	3	1898	1520	2090	1220
1	60	B	3	2	1	2590	2050	2874	936
1	60	B	3	2	2	2530	1978	2862	934
1	60	B	3	2	3	2486	2000	2838	924
1	60	C	1	1	1	3344	4186	3394	4256
1	60	C	1	1	2	3008	3826	2966	3764
1	60	C	1	1	3	2682	3412	2646	3240
1	60	C	1	2	1	4610	6102	4652	3956
1	60	C	1	2	2	4020	5466	4104	3378
1	60	C	1	2	3	3538	4892	3620	2874
1	60	C	2	1	1	2746	3168	2886	3068
1	60	C	2	1	2	2556	2908	2632	2754
1	60	C	2	1	3	2138	2568	2304	2216
1	60	C	2	2	1	3940	4728	3978	2652
1	60	C	2	2	2	3398	4320	3636	2492
1	60	C	2	2	3	2966	3784	3038	2050
1	60	C	3	1	1	1754	1450	1996	1122
1	60	C	3	1	2	1464	1348	1772	1004

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
1	60	C	3	1	3	1338	1312	1586	920
1	60	C	3	2	1	2346	2058	2820	960
1	60	C	3	2	2	2150	1914	2492	814
1	60	C	3	2	3	1838	1812	2138	768
2	40	A	1	1	1	3724	4912	3702	4446
2	40	A	1	1	2	4332	5526	4200	5236
2	40	A	1	1	3	5288	6482	5176	6286
2	40	A	1	2	1	2760	3626	2904	3396
2	40	A	1	2	2	3052	4214	3362	3960
2	40	A	1	2	3	3582	4824	3840	4660
2	40	A	2	1	1	3110	3690	3194	3168
2	40	A	2	1	2	3446	3970	3580	3568
2	40	A	2	1	3	4200	4924	4460	4494
2	40	A	2	2	1	2286	2782	2544	2450
2	40	A	2	2	2	2516	3136	2948	2894
2	40	A	2	2	3	2852	3490	3342	3196
2	40	A	3	1	1	1880	1742	2246	1114
2	40	A	3	1	2	2072	1762	2516	1210
2	40	A	3	1	3	2652	2152	3168	1396
2	40	A	3	2	1	1570	1514	1962	1088
2	40	A	3	2	2	1616	1506	2152	1094

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
2	40	A	3	2	3	1792	1604	2398	1238
2	40	B	1	1	1	3866	4928	3610	4638
2	40	B	1	1	2	3878	4952	3558	4612
2	40	B	1	1	3	3730	4942	3552	4498
2	40	B	1	2	1	2640	3714	2790	3476
2	40	B	1	2	2	2652	3672	2782	3446
2	40	B	1	2	3	2804	3790	2870	3504
2	40	B	2	1	1	3152	3662	3122	3332
2	40	B	2	1	2	3152	3794	3204	3342
2	40	B	2	1	3	3098	3704	3286	3130
2	40	B	2	2	1	2144	2770	2490	2582
2	40	B	2	2	2	2176	2750	2492	2480
2	40	B	2	2	3	2214	2782	2502	2470
2	40	B	3	1	1	1912	1592	2242	1054
2	40	B	3	1	2	1912	1622	2302	1026
2	40	B	3	1	3	1940	1698	2266	1066
2	40	B	3	2	1	1456	1326	1810	990
2	40	B	3	2	2	1498	1346	1742	1008
2	40	B	3	2	3	1504	1396	1874	982
2	40	C	1	1	1	3910	5050	3764	4724
2	40	C	1	1	2	3268	4166	3090	3756

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
2	40	C	1	1	3	2928	3756	2688	3146
2	40	C	1	2	1	2670	3638	2928	3380
2	40	C	1	2	2	2530	3274	2608	2932
2	40	C	1	2	3	2220	3092	2388	2596
2	40	C	2	1	1	3118	3818	3260	3132
2	40	C	2	1	2	2748	3322	2954	2674
2	40	C	2	1	3	2408	3016	2492	2284
2	40	C	2	2	1	2182	2848	2510	2508
2	40	C	2	2	2	2034	2514	2314	2068
2	40	C	2	2	3	1842	2394	2024	1960
2	40	C	3	1	1	1950	1724	2254	1086
2	40	C	3	1	2	1702	1696	2056	1018
2	40	C	3	1	3	1498	1512	1820	830
2	40	C	3	2	1	1464	1396	1896	996
2	40	C	3	2	2	1374	1352	1776	934
2	40	C	3	2	3	1262	1308	1564	824
2	60	A	1	1	1	3944	4944	3778	4602
2	60	A	1	1	2	4670	5672	4476	5432
2	60	A	1	1	3	5388	6428	5264	6216
2	60	A	1	2	1	2708	3572	2904	3398
2	60	A	1	2	2	3098	4182	3312	3958

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
2	60	A	1	2	3	3788	4954	3986	4806
2	60	A	2	1	1	3212	3632	3238	3192
2	60	A	2	1	2	3852	4252	3960	3916
2	60	A	2	1	3	4476	4782	4524	4468
2	60	A	2	2	1	2332	2880	2728	2596
2	60	A	2	2	2	2534	3100	3020	2918
2	60	A	2	2	3	3046	3598	3466	3350
2	60	A	3	1	1	1946	1672	2310	1118
2	60	A	3	1	2	2212	1730	2696	1208
2	60	A	3	1	3	2760	2094	3264	1432
2	60	A	3	2	1	1474	1370	2052	970
2	60	A	3	2	2	1672	1448	2218	1064
2	60	A	3	2	3	1968	1584	2596	1280
2	60	B	1	1	1	3982	5018	3818	4838
2	60	B	1	1	2	4082	5040	3850	4756
2	60	B	1	1	3	3856	4884	3672	4592
2	60	B	1	2	1	2782	3808	3020	3602
2	60	B	1	2	2	2674	3694	2970	3562
2	60	B	1	2	3	2784	3672	2902	3448
2	60	B	2	1	1	3104	3522	3252	3198
2	60	B	2	1	2	3222	3654	3314	3324

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
2	60	B	2	1	3	3254	3728	3338	3318
2	60	B	2	2	1	2120	2658	2562	2464
2	60	B	2	2	2	2230	2652	2620	2470
2	60	B	2	2	3	2290	2784	2644	2522
2	60	B	3	1	1	2070	1492	2380	1044
2	60	B	3	1	2	2096	1470	2408	1034
2	60	B	3	1	3	2146	1622	2412	1092
2	60	B	3	2	1	1544	1284	1912	984
2	60	B	3	2	2	1480	1234	2006	932
2	60	B	3	2	3	1454	1286	1834	994
2	60	C	1	1	1	3918	4918	3760	4614
2	60	C	1	1	2	3530	4432	3364	4002
2	60	C	1	1	3	3084	3964	2936	3464
2	60	C	1	2	1	2700	3584	2824	3376
2	60	C	1	2	2	2498	3302	2672	3060
2	60	C	1	2	3	2176	2938	2364	2682
2	60	C	2	1	1	3130	3602	3204	3218
2	60	C	2	1	2	2776	3274	2904	2802
2	60	C	2	1	3	2550	3034	2640	2550
2	60	C	2	2	1	2306	2770	2690	2518
2	60	C	2	2	2	2060	2578	2276	2246

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
2	60	C	2	2	3	1840	2362	2100	1990
2	60	C	3	1	1	1950	1654	2390	1152
2	60	C	3	1	2	1730	1510	2042	942
2	60	C	3	1	3	1610	1470	1776	876
2	60	C	3	2	1	1582	1374	2098	1016
2	60	C	3	2	2	1274	1248	1750	926
2	60	C	3	2	3	1300	1266	1610	810
3	40	A	1	1	1	3572	4564	3560	4344
3	40	A	1	1	2	4126	5266	4060	5006
3	40	A	1	1	3	4954	6038	4818	6006
3	40	A	1	2	1	3038	4168	3340	3442
3	40	A	1	2	2	3366	4678	3682	3994
3	40	A	1	2	3	3976	5470	4298	4798
3	40	A	2	1	1	2970	3622	3132	3102
3	40	A	2	1	2	3330	3938	3462	3618
3	40	A	2	1	3	4020	4600	4202	4352
3	40	A	2	2	1	2606	3166	2880	2528
3	40	A	2	2	2	2870	3628	3330	2974
3	40	A	2	2	3	3348	4190	3824	3452
3	40	A	3	1	1	1848	1644	2158	1058
3	40	A	3	1	2	2024	1838	2410	1270

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
3	40	A	3	1	3	2444	2138	2890	1514
3	40	A	3	2	1	1700	1650	2084	1050
3	40	A	3	2	2	1758	1666	2340	1094
3	40	A	3	2	3	2028	1900	2716	1244
3	40	B	1	1	1	3540	4604	3378	4528
3	40	B	1	1	2	3610	4578	3506	4464
3	40	B	1	1	3	3572	4616	3428	4402
3	40	B	1	2	1	3056	4160	3150	3570
3	40	B	1	2	2	2866	4194	3026	3482
3	40	B	1	2	3	2936	3990	3056	3300
3	40	B	2	1	1	2938	3532	3144	3314
3	40	B	2	1	2	2912	3442	2918	3250
3	40	B	2	1	3	3048	3534	3016	3150
3	40	B	2	2	1	2542	3268	2880	2632
3	40	B	2	2	2	2418	3124	2788	2510
3	40	B	2	2	3	2526	3230	2778	2568
3	40	B	3	1	1	1776	1480	2114	1070
3	40	B	3	1	2	1732	1562	2052	1094
3	40	B	3	1	3	1978	1742	2276	1168
3	40	B	3	2	1	1568	1494	2082	990
3	40	B	3	2	2	1540	1464	2088	966

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
3	40	B	3	2	3	1584	1580	2010	986
3	40	C	1	1	1	3666	4684	3550	4348
3	40	C	1	1	2	3292	4072	3064	3756
3	40	C	1	1	3	2894	3734	2816	3190
3	40	C	1	2	1	3068	4214	3386	3512
3	40	C	1	2	2	2802	3782	2956	3110
3	40	C	1	2	3	2506	3412	2630	2698
3	40	C	2	1	1	2974	3492	3094	3124
3	40	C	2	1	2	2716	3266	2802	2736
3	40	C	2	1	3	2298	2824	2404	2282
3	40	C	2	2	1	2654	3260	2872	2560
3	40	C	2	2	2	2336	2924	2510	2104
3	40	C	2	2	3	2110	2670	2322	1910
3	40	C	3	1	1	1848	1688	2130	1068
3	40	C	3	1	2	1638	1564	1956	958
3	40	C	3	1	3	1458	1524	1776	902
3	40	C	3	2	1	1634	1612	2038	1030
3	40	C	3	2	2	1486	1522	1924	904
3	40	C	3	2	3	1458	1526	1726	864
3	60	A	1	1	1	3926	4832	3722	4698
3	60	A	1	1	2	4184	5204	4002	5108

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
3	60	A	1	1	3	5026	6020	4838	5950
3	60	A	1	2	1	3012	4026	3266	3344
3	60	A	1	2	2	3538	4734	3722	4060
3	60	A	1	2	3	4008	5342	4232	4684
3	60	A	2	1	1	2916	3398	3124	3142
3	60	A	2	1	2	3458	3888	3636	3686
3	60	A	2	1	3	4110	4528	4316	4350
3	60	A	2	2	1	2636	3172	2970	2548
3	60	A	2	2	2	2804	3448	3206	2830
3	60	A	2	2	3	3498	4142	3926	3472
3	60	A	3	1	1	1758	1494	2084	1056
3	60	A	3	1	2	2064	1754	2558	1258
3	60	A	3	1	3	2630	1950	3018	1486
3	60	A	3	2	1	1770	1486	2254	1022
3	60	A	3	2	2	1792	1542	2376	1028
3	60	A	3	2	3	2188	1770	2862	1274
3	60	B	1	1	1	3692	4502	3418	4442
3	60	B	1	1	2	3756	4658	3584	4546
3	60	B	1	1	3	3694	4648	3512	4516
3	60	B	1	2	1	3080	4232	3216	3534
3	60	B	1	2	2	3100	4264	3270	3616

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
3	60	B	1	2	3	3096	4176	3244	3452
3	60	B	2	1	1	3024	3446	3178	3256
3	60	B	2	1	2	2978	3386	3134	3204
3	60	B	2	1	3	3052	3434	3070	3204
3	60	B	2	2	1	2528	3152	2884	2516
3	60	B	2	2	2	2574	3228	2920	2582
3	60	B	2	2	3	2666	3250	2890	2574
3	60	B	3	1	1	1988	1510	2236	1128
3	60	B	3	1	2	1902	1516	2220	1070
3	60	B	3	1	3	1904	1544	2224	1062
3	60	B	3	2	1	1672	1426	2254	972
3	60	B	3	2	2	1734	1470	2240	986
3	60	B	3	2	3	1792	1524	2236	974
3	60	C	1	1	1	3742	4612	3602	4482
3	60	C	1	1	2	3238	4104	3110	3902
3	60	C	1	1	3	2880	3664	2758	3292
3	60	C	1	2	1	3000	4166	3256	3444
3	60	C	1	2	2	2722	3752	3014	3100
3	60	C	1	2	3	2458	3352	2612	2678
3	60	C	2	1	1	3078	3464	3262	3144
3	60	C	2	1	2	2614	3084	2738	2720

Table 30--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
3	60	C	2	1	3	2368	2828	2540	2388
3	60	C	2	2	1	2530	3124	2908	2490
3	60	C	2	2	2	2206	2854	2634	2234
3	60	C	2	2	3	1970	2534	2288	1840
3	60	C	3	1	1	1878	1598	2212	1152
3	60	C	3	1	2	1696	1470	1968	944
3	60	C	3	1	3	1486	1414	1708	860
3	60	C	3	2	1	1672	1454	2246	922
3	60	C	3	2	2	1512	1458	2000	934
3	60	C	3	2	3	1360	1418	1708	896

Table 31

Estimated All-Pairs Power: α = .05

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
1	40	A	1	1	1	1177	2979	2191	1873
1	40	A	1	1	2	1460	3608	2692	2401
1	40	A	1	1	3	1811	4229	3243	2947
1	40	A	1	2	1	1549	5346	3879	2004
1	40	A	1	2	2	1920	5974	4470	2450
1	40	A	1	2	3	2389	6951	5336	3004
1	40	A	2	1	1	886	2118	1868	1235
1	40	A	2	1	2	1092	2492	2152	1548
1	40	A	2	1	3	1387	3068	2726	1986
1	40	A	2	2	1	1281	3876	3248	1300
1	40	A	2	2	2	1513	4613	3775	1567
1	40	A	2	2	3	1920	5416	4640	2003
1	40	A	3	1	1	559	979	1308	430
1	40	A	3	1	2	620	1057	1508	463
1	40	A	3	1	3	769	1251	1833	580
1	40	A	3	2	1	745	1802	2191	438
1	40	A	3	2	2	893	2005	2525	483
1	40	A	3	2	3	1121	2305	3116	538
1	40	B	1	1	1	1183	3099	2153	2011
1	40	B	1	1	2	1166	3112	2151	2040

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
1	40	B	1	1	3	1188	3077	2175	1997
1	40	B	1	2	1	1592	5481	3746	2123
1	40	B	1	2	2	1587	5416	3755	2029
1	40	B	1	2	3	1572	5401	3706	2034
1	40	B	2	1	1	969	2275	1880	1385
1	40	B	2	1	2	952	2310	1894	1379
1	40	B	2	1	3	964	2351	1978	1348
1	40	B	2	2	1	1282	3963	3160	1312
1	40	B	2	2	2	1247	4093	3221	1305
1	40	B	2	2	3	1256	4124	3165	1306
1	40	B	3	1	1	531	935	1298	404
1	40	B	3	1	2	547	954	1251	431
1	40	B	3	1	3	549	940	1288	411
1	40	B	3	2	1	758	1655	2156	383
1	40	B	3	2	2	758	1751	2189	403
1	40	B	3	2	3	768	1793	2213	394
1	40	C	1	1	1	1167	3091	2291	1912
1	40	C	1	1	2	972	2599	1931	1596
1	40	C	1	1	3	832	2351	1653	1294
1	40	C	1	2	1	1500	5131	3602	1916
1	40	C	1	2	2	1331	4727	3183	1610

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
1	40	C	1	2	3	1102	4089	2742	1325
1	40	C	2	1	1	920	2217	1873	1263
1	40	C	2	1	2	786	1931	1601	1036
1	40	C	2	1	3	658	1715	1390	892
1	40	C	2	2	1	1226	4045	3281	1325
1	40	C	2	2	2	1027	3455	2694	1039
1	40	C	2	2	3	905	3116	2248	897
1	40	C	3	1	1	533	1028	1255	433
1	40	C	3	1	2	496	983	1157	385
1	40	C	3	1	3	403	874	967	329
1	40	C	3	2	1	738	1747	2086	391
1	40	C	3	2	2	625	1588	1780	363
1	40	C	3	2	3	532	1549	1573	357
1	60	A	1	1	1	1247	3085	2273	1986
1	60	A	1	1	2	1545	3650	2780	2506
1	60	A	1	1	3	1954	4437	3396	3123
1	60	A	1	2	1	1616	5330	3894	1910
1	60	A	1	2	2	2050	6186	4719	2461
1	60	A	1	2	3	2532	7232	5656	3165
1	60	A	2	1	1	975	2216	1951	1349
1	60	A	2	1	2	1158	2538	2251	1629

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
1	60	A	2	1	3	1519	3119	2891	2091
1	60	A	2	2	1	1355	4080	3357	1356
1	60	A	2	2	2	1616	4619	3950	1575
1	60	A	2	2	3	1980	5342	4748	1981
1	60	A	3	1	1	615	957	1424	445
1	60	A	3	1	2	692	1064	1611	508
1	60	A	3	1	3	824	1153	1903	550
1	60	A	3	2	1	778	1583	2190	382
1	60	A	3	2	2	944	1840	2650	468
1	60	A	3	2	3	1189	2097	3227	532
1	60	B	1	1	1	1298	3279	2364	2182
1	60	B	1	1	2	1204	3025	2260	2031
1	60	B	1	1	3	1229	3079	2295	2026
1	60	B	1	2	1	1752	5618	4081	2183
1	60	B	1	2	2	1733	5695	4121	2148
1	60	B	1	2	3	1693	5513	3950	2075
1	60	B	2	1	1	1024	2287	2045	1458
1	60	B	2	1	2	1018	2233	1998	1375
1	60	B	2	1	3	967	2283	1943	1377
1	60	B	2	2	1	1381	4099	3424	1297
1	60	B	2	2	2	1302	4016	3340	1335

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
1	60	B	2	2	3	1341	4042	3357	1294
1	60	B	3	1	1	605	882	1380	407
1	60	B	3	1	2	590	896	1313	405
1	60	B	3	1	3	622	996	1383	488
1	60	B	3	2	1	853	1615	2304	382
1	60	B	3	2	2	812	1599	2298	393
1	60	B	3	2	3	809	1621	2289	384
1	60	C	1	1	1	1248	3069	2416	2002
1	60	C	1	1	2	1075	2718	2016	1717
1	60	C	1	1	3	918	2390	1765	1445
1	60	C	1	2	1	1697	5485	3975	2102
1	60	C	1	2	2	1409	4828	3428	1694
1	60	C	1	2	3	1210	4221	2952	1394
1	60	C	2	1	1	970	2245	1996	1338
1	60	C	2	1	2	891	2026	1791	1205
1	60	C	2	1	3	707	1756	1526	940
1	60	C	2	2	1	1376	4112	3356	1345
1	60	C	2	2	2	1168	3705	3009	1200
1	60	C	2	2	3	975	3180	2456	942
1	60	C	3	1	1	574	952	1336	428
1	60	C	3	1	2	484	877	1197	403

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
1	60	C	3	1	3	421	862	1043	368
1	60	C	3	2	1	757	1656	2243	402
1	60	C	3	2	2	673	1526	1963	325
1	60	C	3	2	3	566	1447	1687	317
2	40	A	1	1	1	1022	2713	1953	1561
2	40	A	1	1	2	1226	3132	2261	1896
2	40	A	1	1	3	1583	3715	2786	2341
2	40	A	1	2	1	841	2432	1870	1364
2	40	A	1	2	2	984	2867	2188	1679
2	40	A	1	2	3	1219	3331	2554	2048
2	40	A	2	1	1	811	2000	1638	1056
2	40	A	2	1	2	941	2138	1885	1217
2	40	A	2	1	3	1183	2718	2405	1590
2	40	A	2	2	1	680	1856	1619	959
2	40	A	2	2	2	789	2109	1936	1150
2	40	A	2	2	3	941	2352	2172	1318
2	40	A	3	1	1	478	885	1150	345
2	40	A	3	1	2	529	919	1299	380
2	40	A	3	1	3	687	1120	1630	444
2	40	A	3	2	1	431	945	1216	380
2	40	A	3	2	2	448	974	1316	397

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
2	40	A	3	2	3	509	997	1490	433
2	40	B	1	1	1	1049	2699	1883	1637
2	40	B	1	1	2	1076	2758	1854	1646
2	40	B	1	1	3	1032	2750	1888	1585
2	40	B	1	2	1	841	2545	1823	1449
2	40	B	1	2	2	832	2491	1818	1419
2	40	B	1	2	3	850	2561	1852	1428
2	40	B	2	1	1	843	1972	1614	1110
2	40	B	2	1	2	843	2026	1611	1137
2	40	B	2	1	3	810	1990	1706	1057
2	40	B	2	2	1	664	1844	1592	1007
2	40	B	2	2	2	637	1836	1583	958
2	40	B	2	2	3	646	1821	1580	959
2	40	B	3	1	1	476	813	1132	324
2	40	B	3	1	2	482	830	1193	319
2	40	B	3	1	3	489	882	1148	338
2	40	B	3	2	1	403	863	1108	354
2	40	B	3	2	2	393	817	1067	340
2	40	B	3	2	3	416	884	1165	346
2	40	C	1	1	1	1063	2755	1957	1620
2	40	C	1	1	2	862	2266	1606	1268

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
2	40	C	1	1	3	756	2013	1352	1044
2	40	C	1	2	1	834	2459	1889	1370
2	40	C	1	2	2	734	2187	1669	1152
2	40	C	1	2	3	645	2002	1488	979
2	40	C	2	1	1	816	2032	1684	1035
2	40	C	2	1	2	706	1768	1532	875
2	40	C	2	1	3	595	1608	1306	725
2	40	C	2	2	1	636	1848	1582	938
2	40	C	2	2	2	578	1624	1465	768
2	40	C	2	2	3	512	1552	1241	714
2	40	C	3	1	1	490	891	1129	348
2	40	C	3	1	2	417	891	1057	307
2	40	C	3	1	3	350	742	893	252
2	40	C	3	2	1	398	877	1122	347
2	40	C	3	2	2	365	833	1050	307
2	40	C	3	2	3	321	805	924	276
2	60	A	1	1	1	1098	2718	2004	1609
2	60	A	1	1	2	1347	3209	2394	2007
2	60	A	1	1	3	1626	3745	2893	2381
2	60	A	1	2	1	867	2446	1899	1404
2	60	A	1	2	2	1028	2871	2184	1704

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
2	60	A	1	2	3	1304	3443	2690	2150
2	60	A	2	1	1	866	2018	1722	1086
2	60	A	2	1	2	1081	2348	2126	1371
2	60	A	2	1	3	1290	2656	2444	1609
2	60	A	2	2	1	694	1907	1721	1005
2	60	A	2	2	2	806	2076	1952	1162
2	60	A	2	2	3	1002	2434	2285	1393
2	60	A	3	1	1	495	869	1196	348
2	60	A	3	1	2	566	904	1416	383
2	60	A	3	1	3	739	1100	1719	459
2	60	A	3	2	1	418	858	1289	344
2	60	A	3	2	2	473	912	1383	362
2	60	A	3	2	3	586	1011	1629	457
2	60	B	1	1	1	1135	2744	2032	1719
2	60	B	1	1	2	1138	2795	2077	1717
2	60	B	1	1	3	1077	2684	1968	1626
2	60	B	1	2	1	895	2604	1990	1527
2	60	B	1	2	2	846	2522	1932	1465
2	60	B	1	2	3	867	2516	1904	1438
2	60	B	2	1	1	845	1909	1723	1080
2	60	B	2	1	2	882	1980	1750	1130

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
2	60	B	2	1	3	865	1981	1747	1124
2	60	B	2	2	1	652	1764	1625	968
2	60	B	2	2	2	668	1776	1695	978
2	60	B	2	2	3	685	1855	1673	981
2	60	B	3	1	1	519	759	1222	315
2	60	B	3	1	2	528	759	1224	311
2	60	B	3	1	3	563	858	1252	347
2	60	B	3	2	1	443	795	1182	347
2	60	B	3	2	2	413	784	1273	318
2	60	B	3	2	3	401	801	1129	335
2	60	C	1	1	1	1104	2756	2005	1637
2	60	C	1	1	2	940	2421	1803	1365
2	60	C	1	1	3	816	2147	1529	1193
2	60	C	1	2	1	867	2382	1830	1386
2	60	C	1	2	2	760	2208	1738	1228
2	60	C	1	2	3	625	1940	1488	1028
2	60	C	2	1	1	829	1919	1670	1084
2	60	C	2	1	2	728	1716	1510	934
2	60	C	2	1	3	656	1601	1359	828
2	60	C	2	2	1	694	1869	1737	999
2	60	C	2	2	2	583	1647	1434	834

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
2	60	C	2	2	3	516	1538	1314	734
2	60	C	3	1	1	494	856	1232	358
2	60	C	3	1	2	420	777	1005	281
2	60	C	3	1	3	389	765	911	259
2	60	C	3	2	1	434	892	1309	372
2	60	C	3	2	2	349	782	1062	317
2	60	C	3	2	3	342	770	999	274
3	40	A	1	1	1	826	1553	1178	1295
3	40	A	1	1	2	1028	1860	1348	1591
3	40	A	1	1	3	1272	2202	1659	1973
3	40	A	1	2	1	759	1720	1329	1186
3	40	A	1	2	2	887	1965	1475	1439
3	40	A	1	2	3	1126	2393	1757	1818
3	40	A	2	1	1	666	1180	986	895
3	40	A	2	1	2	770	1295	1126	1063
3	40	A	2	1	3	973	1574	1393	1341
3	40	A	2	2	1	631	1280	1094	828
3	40	A	2	2	2	714	1463	1275	986
3	40	A	2	2	3	904	1740	1508	1222
3	40	A	3	1	1	392	545	709	287
3	40	A	3	1	2	444	604	779	342

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
3	40	A	3	1	3	554	695	941	411
3	40	A	3	2	1	387	628	782	312
3	40	A	3	2	2	397	643	865	322
3	40	A	3	2	3	500	734	1008	377
3	40	B	1	1	1	840	1582	1088	1393
3	40	B	1	1	2	866	1589	1138	1388
3	40	B	1	1	3	841	1594	1137	1348
3	40	B	1	2	1	778	1718	1210	1266
3	40	B	1	2	2	739	1700	1162	1200
3	40	B	1	2	3	714	1648	1174	1120
3	40	B	2	1	1	682	1177	1005	973
3	40	B	2	1	2	663	1135	934	937
3	40	B	2	1	3	699	1192	992	936
3	40	B	2	2	1	634	1309	1100	893
3	40	B	2	2	2	592	1250	1064	827
3	40	B	2	2	3	611	1284	1061	852
3	40	B	3	1	1	385	481	665	291
3	40	B	3	1	2	374	509	648	292
3	40	B	3	1	3	425	567	726	313
3	40	B	3	2	1	361	561	767	294
3	40	B	3	2	2	349	549	759	285

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
3	40	B	3	2	3	350	588	721	290
3	40	C	1	1	1	864	1593	1185	1305
3	40	C	1	1	2	750	1399	1029	1094
3	40	C	1	1	3	625	1258	929	909
3	40	C	1	2	1	776	1737	1305	1204
3	40	C	1	2	2	675	1549	1149	1036
3	40	C	1	2	3	583	1370	1010	877
3	40	C	2	1	1	669	1158	1012	900
3	40	C	2	1	2	601	1083	926	780
3	40	C	2	1	3	488	936	768	632
3	40	C	2	2	1	648	1293	1104	844
3	40	C	2	2	2	542	1158	957	677
3	40	C	2	2	3	463	1031	852	582
3	40	C	3	1	1	384	553	683	291
3	40	C	3	1	2	342	515	625	249
3	40	C	3	1	3	289	510	574	237
3	40	C	3	2	1	360	609	736	296
3	40	C	3	2	2	320	559	684	253
3	40	C	3	2	3	302	581	626	242
3	60	A	1	1	1	934	1651	1204	1423
3	60	A	1	1	2	1060	1826	1337	1622

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
3	60	A	1	1	3	1270	2161	1676	1962
3	60	A	1	2	1	781	1638	1276	1162
3	60	A	1	2	2	961	1987	1492	1486
3	60	A	1	2	3	1137	2320	1743	1778
3	60	A	2	1	1	674	1119	1026	922
3	60	A	2	1	2	823	1292	1198	1093
3	60	A	2	1	3	1016	1537	1424	1333
3	60	A	2	2	1	656	1259	1153	855
3	60	A	2	2	2	713	1388	1237	959
3	60	A	2	2	3	937	1692	1562	1221
3	60	A	3	1	1	380	487	654	281
3	60	A	3	1	2	460	565	823	341
3	60	A	3	1	3	600	642	984	411
3	60	A	3	2	1	400	566	807	296
3	60	A	3	2	2	428	576	878	300
3	60	A	3	2	3	535	671	1082	385
3	60	B	1	1	1	899	1553	1139	1387
3	60	B	1	1	2	901	1602	1193	1414
3	60	B	1	1	3	865	1561	1145	1359
3	60	B	1	2	1	803	1740	1254	1262
3	60	B	1	2	2	810	1746	1251	1298

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
3	60	B	1	2	3	795	1719	1252	1221
3	60	B	2	1	1	704	1111	1006	942
3	60	B	2	1	2	693	1134	1009	966
3	60	B	2	1	3	708	1123	986	935
3	60	B	2	2	1	636	1234	1096	844
3	60	B	2	2	2	654	1308	1118	881
3	60	B	2	2	3	653	1302	1104	868
3	60	B	3	1	1	451	490	724	311
3	60	B	3	1	2	415	471	703	280
3	60	B	3	1	3	416	496	722	290
3	60	B	3	2	1	399	530	822	288
3	60	B	3	2	2	410	550	818	290
3	60	B	3	2	3	407	567	812	288
3	60	C	1	1	1	909	1581	1205	1377
3	60	C	1	1	2	762	1386	1015	1158
3	60	C	1	1	3	650	1217	912	955
3	60	C	1	2	1	758	1705	1269	1183
3	60	C	1	2	2	694	1509	1165	1037
3	60	C	1	2	3	590	1348	1004	882
3	60	C	2	1	1	730	1140	1067	940
3	60	C	2	1	2	580	1017	887	781

Table 31--continued.

CM	N	n_1/n_2	D	ε	$\Sigma_1:\Sigma_2$	B_{DT}	W	W_{RT}	S(C)
3	60	C	2	1	3	513	923	803	672
3	60	C	2	2	1	622	1235	1116	827
3	60	C	2	2	2	537	1128	1000	713
3	60	C	2	2	3	449	975	849	578
3	60	C	3	1	1	411	507	712	301
3	60	C	3	1	2	364	483	627	256
3	60	C	3	1	3	319	476	560	234
3	60	C	3	2	1	387	550	811	274
3	60	C	3	2	2	349	552	734	282
3	60	C	3	2	3	292	536	615	253

REFERENCES

Algina, J. (1994). Some alternative approximate tests for a split plot design. *Multivariate Behavioral Research, 29(4)*, 365-384.

Algina, J. J., & Oshima, T. C. (1994). Type I error rates for Huynh's general approximation and improved general approximation tests. *British Journal of Mathematical and Statistical Psychology, 47*, 151-165.

Algina, J. J., & Oshima, T. C. (in press). An improved general approximation test for the main effect in a split-plot design. *British Journal of Mathematical and Statistical Psychology.*

Berenson, M. L. (1983). Is the F-test really robust with respect to nonnormality?. *ASA Proceeding of Statistical Education Section*, 133-135.

Boik, R. J. (1981). A priori tests in repeated measures designs: Effects of nonsphericity. *Psychometrika, 46*, 241-255.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics, 25*, 290-302.

Bradley, J. C. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.

Brogan, D. R., & Kutner, M. H. (1980). Comparative analysis of pretest-posttest research designs. *The American Statistician, 34(4)*, 229-232.

Cochran, W. G. (1964). Approximate significance levels of the Behrens-Fisher test. *Biometrics, 20*, 191-195.

Collier, R. O. Jr., Baker, F. B., Mandeville, G. K., & Hayes, T. F. (1967). Estimates of test size for several test procedures based on conventional variance rations in the repeated measures design. *Psychometrika, 32*, 339-353.

David, F. N., & Johnson, N. L. (1951). The effect of non-normality on the power function of the F-test in the analysis of variance. *Biometrika, 38*, 43-57.

Donaldson, T. S. (1968). Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio. *Journal of the American Statistical Association, 63*, 660-676.

Dunnett, C. W. (1980). Pairwise multiple comparisons in the unequal variance case. *Journal of the American Statistics Association, 75*, 796-800.

Dunnett, C. W. (1982). Robust multiple comparisons. *Commununications in Statistics-Theory and Method, 11(22)*, 2611-2629.

Gayen, A. K. (1950). The distribution of the variance ratio in random samples of any size drawn from non-normal universes. *Biometrika, 37*, 236-255.

Geisser, S., & Greenhouse, S. W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. *Annals of Mathematical Statistics, 29*, 885-891.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112.

Hack, H. R. B. (1958). An empirical investigation into the distribution of the F-ratio in samples from two non-normal populations. *Biometrika, 45*, 260-265.

Hayter, A. J. (1986). The maximum familywise error rate of Fisher's least significant difference test. *Journal of the American Statistical Association, 81*, 1000-1004.

Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distributions. In D. Hoaglin, F. Mosteller, & J. Tukey (Eds), *Exploring data tables, trends, and shapes*. New York: Wiley.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika, 75*, 800-802.

Huynh, H. (1978). Some approximate tests for repeated measurement designs. *Psychometrika, 43(2)*, 161-175.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean squared rations in repeated measurements designs have exact F distributions. *Journal of the American Statistical Association, 65*, 1582-1589.

Huynh, H., & Feldt, L. S. (1980). Performance of traditional F tests in repeated measures designs under covariance heterogeneity. *Communications in Statistics-Theory and Method, 9(1)*, 61-74.

Imhof, J. P. (1962). Testing the hypothesis of fixed main-effects in Scheffe's mixed model. *Annals of Mathematical Statistics, 33*, 1085-1095.

Keselman, H. J. (1994). Stepwise and simultaneous multiple comparison procedures of repeated measures' means. *Journal of Educational Statistics, 19(2)*, 127-162.

Keselman, H. J., & Keselman, J. C. (1988). Repeated measures multiple comparison procedures: Effects of violating multisample sphericity in unbalanced designs. *Journal of Educational Statistics, 13(3)*, 215-226.

Keselman, J. C., & Keselman, H. J. (1990). Analyzing unbalanced repeated measures designs. *British Journal of Mathematical and Statistical Psychology, 43*, 265-282.

Keselman, H. J., Keselman, J. C., & J. P. Shaffer. (1991). Multiple pairwise comparisons of repeated measures means under violation of multisample sphericity. *Psychological Bulletin, 110(1)*, 162-170.

Lecoutre, B. (1991). A correction for the ε approximate test in repeated measures designs with two or more independent groups. *Journal of Educational Statistics, 16*, 371-372.

MacGillivray, H. L. (1992). Shape properties of the g-and-h and Johnson families. *Communications in Statistics-Theory and Method, 21(5)*, 1233-1250.

Martinez, J., & Iglewicz, B. (1984). Some properties of the Tukey g and h family of distributions. *Communications in Statistics-Theory and Method, 3(3)*, 353-369.

Maxwell, S. E. (1980). Pairwise multiple comparisons in repeated measures designs. *Journal of Educational Statistics, 5(3)*, 269-287.

Maxwell, S. E., & Arvey, R. D. (1982). Small sample profile analysis with many variables. *Psychological Bulletin, 92*, 778-785.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156-166.

Mitzel, H. C., & Games, P. A. (1981). Circularity and multiple comparisons in repeated measures designs. *British Journal of Mathematical and Statistical Psychology, 34*, 253-259.

Myers, J. L. (1979). *Fundamentals of experimental design* (3rd ed). Boston: Allyn and Bacon.

Pearson, E. S. (1931). The analysis of variance in cases of non-normal variation. *Biometrika, 23*, 114-133.

Peritz, E. (1970). A note on multiple comparisons. *Unpublished manuscript*, Hebrew University, Israel.

Quintana, S. M., & Maxwell, S. E. (1985). A *better-than-average estimate of ε.* Paper presented at the Annual Meeting of the American Educational Research Association.

Rogan, J. C., Keselman, H. J., & Mendoza, J. L. (1979). Analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology, 32*, 269-286.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika, 6*, 309-316.

Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of American Statistical Association, 81*, 826-831.

Srivastava, A. B. L. (1959). Effect of non-normality on the power of the analysis of variance test. *Biometrika, 46*, 114-122.

Stoloff, P. H. (1970). Correcting for heterogeneity of covariance for repeated measures designs of the analysis of variance. *Educational and Psychological Measurement, 30*, 909-924.

Tiku, M. L. (1964). Approximating the general non-normal variance-ratio sampling distributions. *Biometrika, 51(2)*, 83-95.

Tiku, M. L. (1971). Power function of the F-test under non-normal situations. *Journal of the American Statistical Association, 66*, 913-916.

Tiku, M. L. (1967). Estimating the mean and standard deviation from censored normal sample. *Biometrika, 54*, 155-165.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In I. Oklin, W. Hoeffding, S. Ghurye, W. Madow & H. Mann (Eds.), *Contributions to probability and statistics*. Stanford, CA: Stanford University Press.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA, Addison-Wesley.

Tukey, J. W., & McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization 1. *Sankyha A, 25*, 331-352.

Welsch, R. E. (1977a). Stepwise multiple comparison procedures. *Journal of the American Statistical Association, 72*, 566-575.

Welsch, R. E. (1977b). *Tables for stepwise multiple comparison procedures* (Working paper No. 949-97). Cambridge: Massachusetts Institute of Technology.

Wilcox, R. R. (1990). Comparing the means of two independent groups. *Biometrical Journal, 32*, 771-780.

Wilcox, R. R. (1992). Comparing one-step M-estimators of location corresponding to two independent groups. *Psychometrika, 57(1)*, 141-154.

Wilcox, R. R. (1993). Analyzing repeated measures or randomized block designs using trimmed means. *British Journal of Mathematical and Statistical Psychology, 46*, 63-76.

Wilson, K. (1975). The sampling distributions of conventional, conservative and corrected F-ratios in repeated measurement designs with heterogeneity of covariance. *Journal of Statistical Computing and Simulation, 3*, 201-215.

Yuen, K. K. (1974). The two sample trimmed t for unequal population variances. *Biometrika, 61*, 165-170.

Yuen, K. K. (1975). Power comparisons of single-sample Winsorized t, trimmed t and Student's t. *The Canadian Journal of Statistics, 3(1)*, 71-80.

Yuen, K. K., & Dixon, W. J. (1973). The approximate behavior and performance of the two-sample trimmed t. *Biometrika, 60*, 369-374.

BIOGRAPHICAL SKETCH

Hyunchul Kim was born February 21, 1960. He received his bachelor's degree with a major in statistics on February, 1983. He next received his master's degree in economics with a major in applied statistics on February, 1985. In the spring of 1992, he began studying for the Ph.D. degree in the Foundations of Education Department at the University of Florida, majoring in evaluation and research methodology. He will graduate with the Ph.D. degree in May, 1995.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James Algina
James J. Algina, Chair
Professor of Foundations
of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Linda M Crocker
Linda M. Crocker
Professor of Foundations
of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

M David Miller
M. David Miller
Associate Professor of
Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Alan Agresti.
Allan Agresti
Professor of Statistics

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 1995

James Algina
Chair, Foundations of
Education

Roderick J. McDavis
Dean, College of Education

Dean, Graduate School